Springer Series in Wireless Technology

Sabina Jeschke
Christian Brecher
Houbing Song
Danda B. Rawat *Editors*

Industrial Internet of Things

Cybermanufacturing Systems

Springer

Springer Series in Wireless Technology

Series editor

Ramjee Prasad, Aalborg, Denmark

Springer Series in Wireless Technology explores the cutting edge of mobile telecommunications technologies. The series includes monographs and review volumes as well as textbooks for advanced and graduate students. The books in the series will be of interest also to professionals working in the telecommunications and computing industries. Under the guidance of its editor, Professor Ramjee Prasad of the Center for TeleInFrastruktur (CTIF), Aalborg University, the series will publish books of the highest quality and topical interest in wireless communications.

More information about this series at http://www.springer.com/series/14020

Sabina Jeschke · Christian Brecher
Houbing Song · Danda B. Rawat
Editors

Industrial Internet of Things

Cybermanufacturing Systems

Springer

Editors
Sabina Jeschke
RWTH Aachen University
Aachen
Germany

Houbing Song
West Virginia University
Montgomery, WV
USA

Christian Brecher
RWTH Aachen University
Aachen
Germany

Danda B. Rawat
Howard University
Washington, DC
USA

ISSN 2365-4139 ISSN 2365-4147 (electronic)
Springer Series in Wireless Technology
ISBN 978-3-319-42558-0 ISBN 978-3-319-42559-7 (eBook)
DOI 10.1007/978-3-319-42559-7

Library of Congress Control Number: 2016948801

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Cyber-Physical Systems for Production Technology

It seems world-famed engineer and inventor Nikola Tesla already predicted the mobile phone about a hundred years ago when he said:

> When wireless is perfectly applied, the whole earth will be converted into a huge brain, which in fact it is, all things being particles of a real and rhythmic whole. A man will be able to carry one in his vest pocket. (Nikola Tesla, 1926)

He had already foreseen that if we could gather all the information in the world, we would indeed get very different insights on how processes are running. And this is exactly the vision of the Internet of Things (IOT) and cyber-physical systems (CPS): Networking everything to facilitate access and enhance performance. The term “cyber-physical system” emerged around 2006, when it was coined by Helen Gill at the National Science Foundation in the USA. She associated the term “cyber” to such systems, which are used for discrete processing and communication of information, while with “physical” the natural man-made technical systems are meant which operate continuously.

> Cyber-physical systems are physical, biological, and engineered systems whose operations are integrated, monitored, and/or controlled by a computational core. Components are networked at every scale. Computing is deeply embedded into every physical component, possibly even into materials. The computational core is an embedded system, usually demands real-time response, and is most often distributed. (Helen Gill, April 2006)

According to Gill, CPSs are therefore systems where virtual and real systems are linked closely at various levels and the components are networked at every scale. As an intellectual challenge, CPS is about the *intersection*, not the union, of the physical world and the cyberspace.

However, the roots of the term CPS are older and go deeper. It would be more accurate to view the terms “cyberspace” and “cyber-physical systems” as stemming from the same root “cybernetics,” rather than viewing one as being derived from the other. The term “cybernetics” was coined by Norbert Wiener in 1948.

Wiener—an US mathematician and later Nobel Laureate—had a huge impact on the development of control systems theory. He described his vision of cybernetics as the conjunction of control and communication. His notion of control was deeply rooted in closed-loop feedback, where the control logic is driven by measurements of physical processes, and in turn drives the physical processes. Even though Wiener did not use digital computers, the control logic is effectively a computation, and therefore, cybernetics is the conjunction of physical processes, computation, and communication.

In the early nineties, US computer scientist Mark Weiser became well known for his concept of "ubiquitous computing." He refers to the perception of a comprehensive computerization and networking of the world and its many objects. Weiser paid early attention to the behavioral changes that occur when the environment is permeated by digital technologies and computing is made to appear anytime and everywhere. According to his vision, computers will disappear as a single device and will be replaced by "intelligent objects." To date, computers and the Internet are the subject of human attention. The so-called Internet of Things should imperceptibly support people in their activities with ever getting smaller computers, without distracting them or even get noticed.

This brings us to the differences of the Internet of Things and cyber-physical systems. Today, they are more or less synonym. The frontier between CPS and IOT has not been clearly identified since both concepts have been driven in parallel from two independent communities, although they have always been closely related. The US scientists at first used the term "Internet of Things" in 1999, more specifically Kevin Ashton, at that time an employee at Procter & Gamble. On June 22, 2009, he wrote in the RFID Journal:

> If we had computers that knew everything there was to know about things—using data they gathered without any help from us—we would be able to track and count everything, and greatly reduce waste, loss and cost. We would know when things needed replacing, repairing or recalling, and whether they were fresh or past their best. (Kevin Ashton, June 2009)

The IOT represents a major extension of the classic Internet: While the Internet is limited to the exchange of data and documents of various media types, the IOT addresses networking with everyday objects. The physical and digital world is merging. In other words, the intelligence is "embedded": Systems gain some kind of intelligence, such as cooperating robots, intelligent infrastructures, or autonomous and interconnected cars. They have certain skills to perceive their environment and communicate with each other, typically via Internet protocols. Thus, "things" are able to communicate.

This is the vision of these two great concepts—IOT and CPS—and the terms are in fact mostly interchangeable as long as we discuss their technological basis. However, the mind-set of the two concepts originates from two different communities: IOT is driven by computer sciences and Internet technologies, it understands itself as an extension of the Internet concept, and it focuses on openness and networks. CPS is driven by engineering aspects and concentrates on the physical systems behind, often in a closed-loop system, which now should start to communicate and cooperate with each other. This difference may be hairsplitting, but it causes huge differences in the methods applied to understand these upcoming systems. In particular, they lead to different modeling, control, and steering paradigms.

In this context, the term "Industry 4.0" was first used in 2011 at the Hannover Fair in Germany. It embraces a number of contemporary automation, data exchange, and manufacturing technologies and has been defined as follows:

> […] a collective term for technologies and concepts of value chain organizations which draws together cyber-physical systems in first article (p. 17 (p3)), the Internet of Things and the Internet of Services. (Wikipedia on Industry 4.0, May 2016)

Industry 4.0 comprises the fourth industrial revolution driven by the Internet. It describes technological changes from today's production technology to cyber-physical production systems. Production machineries such as welding robots, conveyor belts, or transportation robots "talk" to each other and cooperate which ultimately leads to an intelligent smart factory.

Keeping in mind that research and developments on IOT and CPS are still in their infancies, the editors have compiled a book to address certain perspectives on specific technological aspects, such as communication networks for cyber-physical systems, today's applications and future potential of cyber-physical systems for agricultural and construction machinery, or approaches from the field of Machine Learning and Big Data for the Smart Factory.

The idea of this book is to use the opportunities coming along with the digitalization and modern networking technologies to record and promote the fourth industrial revolution in the area of production technology and related fields. The book documents the first steps of this revolution with a broad selection of different authors and provides food for thought for the next steps. These networking technologies are not limited to certain areas, but address broad areas of our society. Therefore, the editors asked different authors to comment on specific issues, such as today's application and future potential of CPS for agricultural and construction machinery or within wind energy or the impacts of CPS for competence management.

It is a technological book with interdisciplinary extensions, just because 4.0 will change everything but will happen with completely different approaches. It is time to deal intensively with questions of how we intend to exploit this enormous potential. Which player will be seen in future on the market? Which jobs have a future? What types and which nations lead the innovation? What does the computer intelligence mean for business models?

I am impressed by the interdisciplinary nature and the high scientific level of this book: The international composition of these 27 scientific contributions of US and European authors is quite outstanding. On the one hand, those two groups agree very closely on several of their views on CPS, but on the other hand, there are different mind-sets driven from different nationalities. Therefore, this collection is an attempt to close the "gap." The variety of articles gives excellent insights, and I hope that the reader will gain as many ideas and inspiration for their research as I did.

Prof. Dr.-Ing. Dr. h.c. Peter Göhner
Former director of the Institute of Industrial Automation and
Software Engineering at the University of Stuttgart

Contents

Author Biography

Danda B. Rawat is an Associate Professor in the Department of Electrical Engineering & Computer Science at Howard University, Washington, DC, USA. Prior to Howard University, he was with the College of Engineering & Information Technology of Georgia Southern University, Statesboro, GA as a faculty member until 2016. Dr. Rawat's research focuses on wireless communication networks, cyber security, cyber physical systems, Internet of Things, big data analytics, wireless virtualization, software-defined networks, smart grid systems, wireless sensor networks, and vehicular/wireless ad-hoc networks. His research is supported by US National Science Foundation, University Sponsored Programs and Center for Sustainability grants. Dr. Rawat is the recipient of NSF Faculty Early Career Development (CAREER) Award. Dr. Rawat has published over 120 scientific/technical articles and 8 books. He has been serving as an Editor/Guest Editor for over 10 international journals. He serves as a Web-Chair for IEEE INFOCOM 2016/2017, served as a Student Travel Grant Co-chair of IEEE INFOCOM 2015, Track Chair for Wireless Networking and Mobility of IEEE CCNC 2016, Track Chair for Communications Network and Protocols of IEEE AINA 2015, and so on. He served as a program chair, general chair, and session chair for numerous international conferences and workshops, and served as a technical program committee (TPC) member for several international conferences including IEEE INFOCOM, IEEE GLOBECOM, IEEE CCNC, IEEE GreenCom, IEEE AINA, IEEE ICC, IEEE WCNC and IEEE VTC conferences. He is the recipient of Outstanding Research Faculty Award (Award for Excellence in Scholarly Activity) 2015, Allen E. Paulson College of Engineering and Technology, GSU among others. He is the Founder and Director of the Cyber-security and Wireless Networking Innovations (CWiNs) Research Lab. He received the Ph.D. in Electrical and Computer Engineering from Old Dominion University, Norfolk, Virginia. Dr. Rawat is a Senior Member of IEEE and member of ACM and ASEE. He served as a Vice Chair of the Executive Committee of the IEEE Savannah Section and Webmaster for the section from 2013 to 2017.

Part I
Introduction and Overview

Industrial Internet of Things and Cyber Manufacturing Systems

Sabina Jeschke, Christian Brecher, Tobias Meisen, Denis Özdemir and Tim Eschert

1 Introduction

The Internet of Things (IoT) is an information network of physical objects (sensors, machines, cars, buildings, and other items) that allows interaction and cooperation of these objects to reach common goals [2]. Applications include among others transportation, healthcare, smart homes and industrial environments [28]. For the latter, the term Industrial Internet of Things (IIoT) or just Industrial Internet is typically used, see e.g. [12]. In this book we will use IIoT synonymously to Industry 4.0 or to the original German term "Industrie 4.0". The differences between the terms or initiatives mainly concern stakeholders, geographical focus and representation [3]. Further, IIoT semantically describes a technology movement, while Industry 4.0 is associated with the expected economic impact. That is to say, IIoT leads to the Industry 4.0. But considering both as research and innovation initiatives, one will not find any technology that is claimed by only one of these. For the title, however, we chose IIoT, because it highlights the idea of networks, which is a cornerstone of many contributions in this book. Further, this book can be regarded as a manufacturing-oriented extension to our collected edition on cyber-physical systems

S. Jeschke · T. Meisen
IMA/ZLW & IfU, RWTH Aachen University, Dennewartstr. 27, 52068 Aachen, Germany
e-mail: sabina.jeschke@ima-zlw-ifu.rwth-aachen.de

T. Meisen
e-mail: tobias.meisen@ima-zlw-ifu.rwth-aachen.de

C. Brecher · D. Özdemir (✉) · T. Eschert
Laboratory for Machine Tools and Production Engineering (WZL),
RWTH Aachen University, Steinbachstrasse 19, 52074 Aachen, Germany
e-mail: D.Oezdemir@wzl.rwth-aachen.de

C. Brecher
e-mail: xcluster@wzl.rwth-aachen.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_1

that contains many foundational topics of IoT [23]. Please note, that in our understanding the IIoT not only is the network of the physical objects in industry but also includes the digital representations of products, processes and factories such as 3D models or physical behavior models of machines.

In the year 2015, IoT has been declared one of the most hyped technologies [11]. Its industrial applications, i.e. IIoT, were even the focus of the World Economic Forum 2016 (*Slogan: Mastering the Fourth Industrial Revolution*). But critical voices are gaining weight. A recent edition of “Handelsblatt” (Germany’s largest business newspaper) that was titled “The efficiency lie” [21] and the new book by the economist Robert Gordon argue that the expected productivity growth from digitalization is small compared to the preceding industrial revolutions are just two examples of this counter movement [14].

In the light of these critical voices it is even more important to analyze where real value can be gained from IIoT in terms of time, flexibility, reliability, cost, and quality. Therefore, we and the other editors are pleased to present many contributions with specific manufacturing applications and use cases in this book. But beyond these concrete scenarios we want to convey the vision of cognitive self-optimizing production networks enabling rapid product innovation, highly individual products and synchronized resource consumption. Therefore, the contributions of this book and the results of the large research initiatives associated with IIoT and Industry 4.0 represent a first step towards these results.

To guide the reader through the book, we will first give a short overview on the history and foundations of IIoT and define the key-terms of this book. Subsequently, the reader may find our overview on global research initiatives helpful for understanding the contributions of this book in the international context. The reader will find slightly different definitions of the key terms throughout the chapters of this book due to these different initiatives. But to give some orientation to the reader, the last part provides a brief summary of the chapters of this book considering the challenges, solutions and forecasts for IIoT.

2 Foundations of the Industrial Internet of Things and Cyber Manufacturing Systems

IIoT has grown from a variety of technologies and their interconnections. In manufacturing, the first attempts to create a network of “things” date back to the 1970s and were summarized with the term “Computer-Integrated Manufacturing” (CIM). Although the ideas of CIM are now approximately 40 years old, most challenges are still prevailing today, e.g. the integration of managerial and engineering processes and the realization of flexible and highly autonomous automation. However, in the 1990s—with the rise of Lean Production—excessive IT solutions were increasingly regarded as inefficient and many CIM projects as a failure. In retrospective, the early disappointments can be traced back to the reason that technology and people were not ready to successfully implement the ideas, e.g.

- Immature IT and communication infrastructure
- Lack of computational power
- Lack of data storage capacity
- Limited connectivity and data transfer rates
- Missing openness of software tools and formats for data exchange.

Moreover, the CIM movement reached its peak before the great breakthrough of the internet between the mid-1990s and the first years of the new millennium. Now, it is difficult to imagine a world without the internet. However, in the 1980s it was difficult to convey the idea of ubiquitous connectivity. In retrospective, it was almost impossible to realize information exchange on a broad scale within the factory at a time when the rest of the world was mostly not digitally connected.

While CIM was focusing on solutions for the shop floor, Product Data Management (PDM) has been established as a new approach to design networks within engineering departments connecting product data and people. In contrast to CIM, PDM was less a technology push, but originated from the limits of handling large amounts of product data with simple file based systems. Functions like product configuration, workflows, revisions, or authorization are now indispensable for engineering departments in large enterprises and are increasingly important for medium-sized companies. With Product Lifecycle Management (PLM) the network idea is taken further, considering consistent data management as an objective for the whole lifecycle [8]. In this context, PDM is usually regarded as the backbone of PLM, providing interfaces to different applications during the lifecycle such as production and service. Therefore, PDM and PLM are also a prerequisite for IIoT: The industrial "things" require product data as a basis for a meaningful communication, e.g. for comparing measurement data to the initially specified requirements associated with the product.

From the perspective of factory planning and operation, the Digital Factory aims to integrate data, models, processes, and software tools [17, 25]. Therefore, the Digital Factory is a comprehensive model of the real factory that can be used for communication, simulation and optimization during its life cycle. Software products in the domain of the Digital Factory typically come with different modules enabling functions such as material flow simulation, robot programming and virtual commissioning. In the context of IIoT, the Digital Factory can be regarded as the complement to PLM. While PLM aims to integrate data along the product life cycle, the Digital Factory comprises the data of production resources and processes. For the IIoT both are necessary, high-fidelity models of the product and its production, see Fig. 1.

While PLM and the Digital Factory contribute to the data backbone of the IIoT, many ideas of designing the hardware for IIoT can be traced back to the idea of mechatronics and Cyber-Physical Systems (CPS). Mechatronics is typically defined as the discipline that integrates mechanics, electronics and information technology [25]. As the term "mechatronics" indicates by its first syllable, the discipline can be regarded as an extension of mechanics and many of the stakeholders have a background in mechanical engineering. In contrast, the name Cyber-Physical

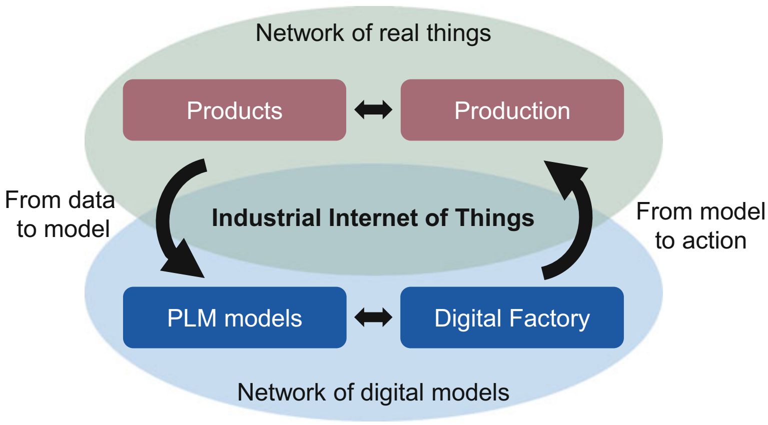


Fig. 1 IIoT as the network of real things and their digital counterparts

Systems has been established by researchers from computer science and software engineering. NASA defines CPS as an "emerging class of physical systems that exhibit complex patterns of behavior due to highly capable embedded software components" [22]. A similar definition is used in the roadmap project CyPhERS: "A CPS consists of computation, communication and control components tightly combined with physical processes of different nature, e.g., mechanical, electrical, and chemical" [6]. The latter definition could also be associated with mechatronic systems and indeed, the terms "mechatronics" and CPS are often used interchangeably, especially in the domains of automation and transport. However, the underlying "engineering philosophy" is usually different. While "mechatronics" implies that there is a physical system in the focus with a software grade-up, CPS indicates that the largest part of added-value is based on software and that the hardware-part is a special challenge for software engineering due to spatiotemporal interaction with the physical environment. Further, a CPS is characterized by the communication between subsystems that is not necessarily part of mechatronics. In this context, the CPS can be characterized as a networked system and usually the network connotation is implicitly included in the term CPS, e.g. by definitions like: CPS comprise "embedded computers and networks [that] monitor and control the physical processes […]" [18]. Taking the network idea further, CPS can be considered as "IoT-enabled" [9], where IoT implies that the subsystems are connected to the internet and therefore part of an open system with a vast number of nodes. Due to their network characteristic, CPS require a larger theoretical foundation than mechatronic systems. While the former can typically be described by the means of multi-physical modeling and control theory, the theory of the latter includes, amongst others, mechatronics, network technology, collaboration methods, cyber security, data analytics, artificial intelligence and human machine interaction. For a summary on the theory and applications of CPS we refer the collected edition of Song et al. [23] and especially to the corresponding introduction by Törngren et al. [24].

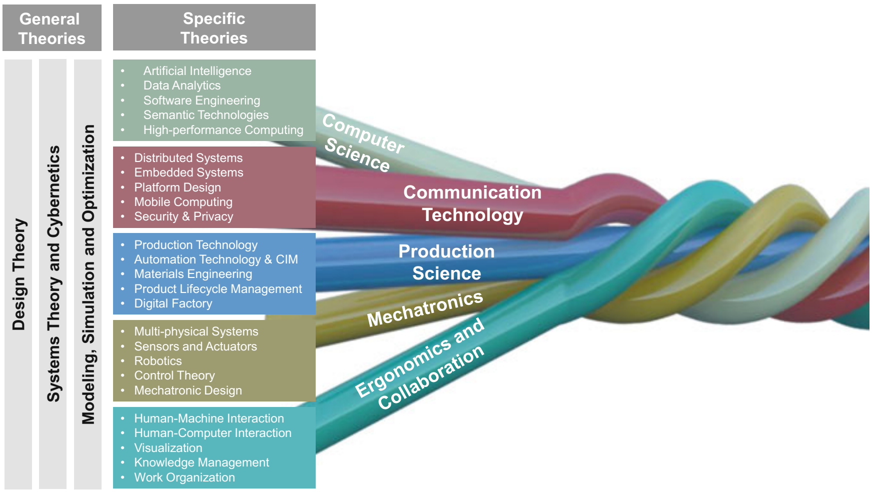


Fig. 2 Theoretical foundations of cyber manufacturing and IIoT

In the context of manufacturing, Cyber Manufacturing Systems (CMS) and IIoT denote the respective industrial counterparts of CPS and IoT. CMS or Cyber-Physical Production Systems (CPPS) are therefore advanced mechatronic production systems that gain their intelligence by their connectivity to the IIoT. Therefore, CMS cannot be considered without IIoT and vice versa. Typically, when one concept is mentioned, the other concept is implicitly included, as in the definition by Lee et al. [19]:

> "Cyber Manufacturing is a transformative concept that involves the translation of data from interconnected systems into predictive and prescriptive operations to achieve resilient performance".

Overall, CMS and IIoT are not individual technologies with a closed theory framework, but rather an interdisciplinary blend from the domains of production, computer science, mechatronics, communication technology and ergonomics, see Fig. 2. Applications of some general theories, however, can be found across all of the disciplines. Systems theory and cybernetics can be seen as the most general approach to describe the interaction between different people and things with the aim to design cybernetic feedback loops that lead to self-optimizing and robust behavior. To understand, predict, and optimize the system behavior it is a common approach to build models that can simulate the system dynamics. Further, system design includes creative action that can generally be put into the framework of design theory, e.g. design thinking. These general theories can be considered as the "glue" for the individual domains that enables to leverage the synergy between them.

3 Potentials and Challenges

Currently, most studies agree that IIoT and CMS as promoted in initiatives such as Industry 4.0 will have a great economic impact. For example, a recent survey by PwC, a consultancy, concludes that the future global cost and efficiency gains by Industry 4.0 will exceed 400 bn. Dollars annually [13]. Countries with a large industry sector such as Germany, where industry has a 30 % share of GDP and employs 25 % of the labor force [4], are challenged by digitalization as the successful transformation to IIoT and CMS is likely to determine the future economic success of the whole economy. This transformation is especially crucial for the sector of machinery and equipment manufacturing as an enabler for other industry sectors. A recent article in the Economist put the challenge in a nutshell by asking "Does Deutschland do digital?", suggesting Germany should withdraw the reservations on platforms and data sharing and should change its corporate culture towards risk-taking and its approach to software engineering towards higher user-friendliness [7].

The transformation to Industry 4.0 is of course no end in itself, but it must lead to greater resource efficiency, shorter time-to-market, higher-value products and new services. More specifically, applications and potential benefits include:

- Intelligent automation that makes small batch sizes down to batch size one feasible because programming and commissioning efforts become negligible
- High-resolution production that improves predictability and cost transparency
- Intelligent production planning that improves the adherence to delivery dates and reduces costs and throughput times
- Predictive maintenance and automatic fault detection leading to a higher overall equipment effectiveness and a reduction of maintenance costs
- Intelligent process control aiming for zero waste, low tooling costs, minimal resource consumption and short running-in and production times
- Reconfigurability that enables quick scale-up and change management
- Human-machine interaction leading to higher labour productivity and improved ergonomics
- Feedback from production to engineering that improves the production systems of the next generation
- Implementation of new business models that leverage the seamless pipeline from customer requirements to product delivery and service

While CPS and IIoT generally have a broad field of application, as shown by the application matrix in Chap. "An Application Map for Industrial Cyber-Physical Systems", the approaches from other fields such as healthcare, transport or energy are not directly transferable. The specific points of CMS and IIoT include:

- Integration from factories to machines and their components
- Life-cycle integration of products and production resources
- Heterogeneous production infrastructure from different suppliers
- Implementation of new systems into systems of existing machinery

- Spatio-temporal relationships between objects in the system
- Broad field of manufacturing technologies
- Humans in versatile operating conditions

Generally, both CMS and IIoT can be regarded as complex systems of systems. Hence, there is not just one technological basis to build such systems, which results in a first challenge: the technological basis and suitable architectures.

A further major challenge is the specification of a generally accepted, extensible infrastructure or architectural pattern that supports, on the one hand, a variety of sensors, actuators, and other hardware and software systems, while on the other hand the complexity of the system has to remain manageable. Such a networked system contains on a small scale a sensor device, but also management or planning systems that give access to enterprise information (e.g. highly aggregated key performance indicators like the overall equipment effectiveness or a bulk of information like the stock of components, parts, and products). In order to manage the various systems and to provide a way to satisfy the information demands, researchers as well as industrials have introduced several pseudo-standardized architectural system patterns in the past. In the field of automation, exemplarily, the well-known automation pyramid or the more advanced automation-diabolo, [27] represent such architectural patterns. With the introduction of CMS and IIoT in automation, these well-structured and task-oriented patterns resolve. As shown in Fig. 3, the classical automation pyramid will be gradually replaced with networked, decentralized organized and (semi-)automated services [26]. Subsequently, new

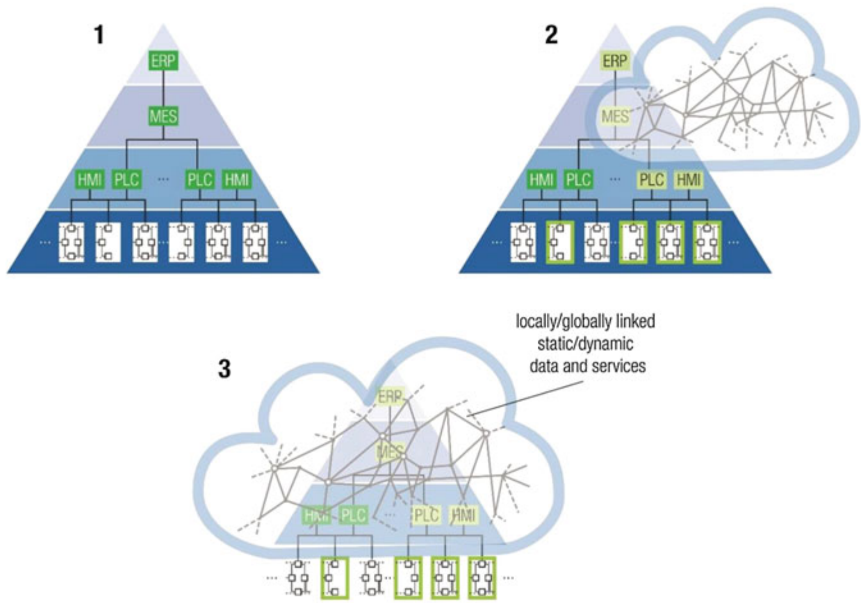


Fig. 3 Gradually replacement of the classical automation pyramid [26]

modeling and design techniques will be required for these networked structures that monitor and control physical production and manufacturing processes.

The evolving infrastructures of CMS and IIoT raise new challenges regarding communication (respectively information exchange). A transparent and adaptive communication is necessary to guarantee real-time delivery of information, robustness and other aspects of Quality-of-Service. Furthermore, such a decentralized system needs a higher level of automation regarding self-management and maintenance. Artificial intelligence and analytics need to be established to facilitate the aforementioned self-management and diagnosis capabilities. Besides, new optimization potentials can be revealed by making use of enormous amounts of gathered data.

Last, human-machine-interfaces have to be adapted reflecting the increasing complexity of these systems. It is necessary that the system ensures a timely and correct display of necessary information. Otherwise, the mass of information cannot be handled by the human worker and decisions cannot be made in time. The manner in which humans interact with the system changes—from human centered control, to an equivalent interaction, in which the cognitive capabilities of the human become central, resulting at least to an evolution of workforce.

4 Major Research Initiatives

To leverage the expected potentials of CMS and IIoT by meeting the aforementioned challenges, major research initiatives have been started across the globe. We want to give a brief summary:

(1) In Germany, major industry associations form the “Plattform Industrie 4.0” that conducts research, advocates for standardization, and coordinates technology transfer and communication between research and industry. Additionally, the topics of CPPS and IIoT are part of major research and innovation projects such as the Leading-Edge Cluster it’s OWL or the Cluster of Excellence “Integrative Production Technology for High-wage Countries” at RWTH Aachen University.

(2) The United States follow a more data-driven approach, mainly led by the “Industrial Internet Consortium (IIC)” and the “National Institute of Standards and Technology (NIST)”, the regulation agency tasked with coordinating the National Network for Manufacturing Innovation (NNMI).

(3) In Japan, most research is taking place in private companies, such as Fanuc or Fujitsu, funded by the “Ministry of Economy, Trade and Industry (METI)”.

(4) The “Ministry of Science and Technology (MoST)” is the coordinator of China’s high-tech strategy. The challenges China currently faces are different from the previously mentioned: Currently, China is a low-wage country, but wages are rising. Environmental pollution is becoming an increasing problem

but not yet fully recognized. Subsequently, technology is approached at high speed and with massive availability of capital lead.

(5) Research in South Korea is mainly driven by the Ministry of Trade, Industry and Energy (MoTIE) and the Ministry of Science, ICT and Future Planning (MSIP), together with one of Korea's largest technical universities: Korea Institute of Industrial Technology (KITECH). Korea is bringing smart manufacturing technologies to implementation, with a focus on safety and under energy constraints, as energy costs are rising.

(6) The "Ministry of Economic Affairs (MOEA)" is responsible for coordinating research in Taiwan. "Speed to market" and "Speed to volume" are the country's two main challenges. Taiwan's Foxconn is the world leader in producing ICT and semiconductors. Moreover, the biggest Taiwanese research institute in the field, The Industrial Technology Research Institute (ITRI), has a competence center for industrial research.

Thus, the way high-tech research is approached in different parts of the world is different and driven by the individual country's needs. However, a field with the global potential of the Internet of Things can only succeed if sharing knowledge and creating global standards become common goals among leaders in politics, research, and industry.

5 Approaches and Solutions

In this section, we give a short overview of the aforementioned grand challenges and the approaches and solutions that are discussed in more detail in the remaining chapters of this book. Thereby, we will extend the list of challenges regarding CMS and IIoT. However, several more technical (like safety and security aspects) as well as non-technical challenges (like suitable business models and the societal impact) exist, but are out of the scope of this book.

5.1 Modeling for CPS and CMS

Model-based design and development of production and manufacturing systems is a crucial task and has been researched for many years. Still, with the rise of CPS new challenges evolve. Nowadays, established models and methods cover e.g. different engineering and software aspects and often impose an early separation between these aspects. Thereby, modeling refers to a formalized approach facilitating the specification of the whole system or parts of it, its behavior as well as its structure. Several modeling tools and tool chains exist from both disciplines engineering as well as computer science. In the context of CPS and CMS, it is necessary to bring these solutions together.

Such integrated tool chains have to cover the different non-functional requirements as for example multidisciplinary and collaboration as well as functional requirements like the realization of hardware and software. Finally, they need to enable analysis, simulation, testing and implementation of the modelled system. Gamble et al. [10] provide an overview as well as deeper insights and discuss ongoing challenges and open research questions in this area.

In this book, modeling of CPS in general and of CMS (or CPPS) in particular is discussed from three different perspectives:

- An overview on CPS engineering for manufacturing is given in Chap. "Cyber-physical Systems Engineering for Manufacturing" from the perspective of the National Institute of Standards and Technology (NIST) in the US. The convergence of different domains poses new and great challenges to standardization tasks. While there is more or less a globally accepted way of mechanical design, there is no such standard for systems engineering. With this background, the article gives an overview on current approaches to system design with special regards to the activities of NIST.
- In Chap. "Model-Based Engineering of Supervisory Controllers for Cyber-Physical Systems" the authors discuss the modeling of supervisory controllers for CPS. Thereby, they describe a supervisory controller as the coordination component of the behavioral aspects of the CPS. Besides highlighting the steps of modeling, supervisory control synthesis, simulation-based validation and visualization, verification, real-time testing, and code generation, the chapter discusses the benefits of the Compositional Interchange Format language in this context.
- Chapter "Formal Verification of SystemC-based Cyber Components" deals with modeling of cyber components. The authors focus on the computation part of CPS—which they summarize as cyber components. Due to the increasing complexity of these components, a modeling on a high level of abstraction is necessary. They provide a new approach that transforms the SystemC model to C and embeds the Transaction Level Modeling (TLM) property in form of assertion into the C model. Furthermore, they present a new induction method for the verification of TLM properties.
- In Chap. "Evaluation Model for Assessment of Cyber-physical Production Systems" the authors examine how CPPS technology can be assessed regarding the value-adds. They give answers to the questions: "How to model the various system characteristics and abilities which are unique to Cyber-Physical Systems?" and "Which indicators and metrics could be utilized to assess the systems performances?" As a result, they provide a model of high level description of Cyber-Physical Technologies.

5.2 Architectural Design Patterns for CMS and IIoT

As pointed out, several pseudo-standardized high-level architectural system patterns exist for production systems. In addition, other domain-specific best practices have emerged over the years. But, with the introduction of CMS, these patterns are questioned. In CPPS, data, services and functions are stored and processed where they are needed and not according to the levels of the automation pyramid [26]. Hence, new design patterns arise, like service-oriented and cloud-based architectures [5, 15, 20]. For such architectures, design patterns, as pre-verified and reusable solution to a common problem in CPS, are yet to be identified and defined. Thereby, especially in the domain of production systems, migration aspects have to be covered.

In this book, such reusable and proven solutions to architectural questions are discussed in the following chapters:

- In Chap. "CPS-based Manufacturing with Semantic Object Memories and Service Orchestration for Industrie 4.0 Applications" the authors present an approach using Virtual Representation (VR). The basic idea relies on the attachment of a virtual representation and a storage space, named the digital object memory, to each physical entity. This digital shadow is furthermore used by actuators and coordination services to orchestrate the production. Furthermore, the chapter discusses additional elements of Industry 4.0 and points out its advantages like "plug'n' produce".
- The aspect of integrating robot-based CPS modules into an existing infrastructure is discussed in Chap. "Integration of a knowledge database and machine vision within a robot-based CPS". Thereby, the chapter covers applications in various industries (e.g. laundry logistics and assembly tasks). Furthermore, the authors reflect on the integration of technologies such as machine vision, RFID and physical human-robot interaction. In doing so, they also explore the possibilities for integration within heterogeneous control systems based on available standards.
- In Chap. "Interoperability in Smart Automation of Cyber Physical Systems" the authors examine interoperability on all levels of automation. They present an approach that is based on semantic technology and standardized, CPS applicable protocols like OPC UA and DDS. Further, they point out use cases, where the technology stack has been successfully used.
- Enhancing the resiliency in production facilities by using CPS, is topic of Chap. "Enhancing resiliency in production facilities through Cyber Physical Systems". Therefore, the authors first review the basic concepts of CPS in factories and their dedicated specificities. By reference to two examples, they further describe the presented concepts in actual facilities.

5.3 Communication and Networking

Humans as well as software and hardware systems produce, procure, distribute, and process data (or, if the needed capabilities are available, information) along a more or less formalized process. Initial objects of this process are data, which are collected, processed, stored, and transmitted with—in case of a technical system involvement—the help of information and communication technology. The final objects of this process are information that the user or another technical system utilize for task fulfilment or to satisfy the need for information (e.g. to make a decision).

In case of CMS, the decentralized communication and the high number of networked participants makes an adaptable and flexible information exchange between the participants necessary. In case new participants are added to the network and others are removed, the information flow still needs to be stable and reliable. In case mandatory information providers are not available, the system needs to react autonomously and accordingly. These requirements necessitate new standardized, extended protocols and network technologies for communication and networking in CPS. Existing concepts have to be analyzed and critically questioned. Semantic technologies, artificial intelligence, and context-awareness are crucial in fulfilling this challenge.

Communication and networking are discussed more detailed in:

- In Chap. "Communication and Networking for the Industrial Internet of Things ", first the characteristics and requirements of CPS are analysed and categorized. Second, the authors map the identified categories to existing communication and networking technologies to discuss the respective technologies in-depth. Thereby, they focus on their applicability to supporting CPS and shortcomings, challenges, and current research efforts.
- A similar analysis is performed in Chap. "Communications for Cyber-Physical Systems". In contradiction to the previous chapter, this one focusses on the communication within CPS in Smart Grids. The authors provide different types of communication networks for CPS that can be encountered at different system levels. They furthermore give an overview of prominent communication standards and protocols adopted in these types of CPS networks and identify open research issues that still need to be addressed.

5.4 Artificial Intelligence and Analytics

The importance of aggregating, processing, and evaluating information increases drastically in IIoT. Enabling the system to self-optimize the workflow and to identify errors and maintenance tasks on its own requires advanced analytic capabilities. Relying on human expertise alone does not work in CPS anymore. Instead, the system has to perform self-optimization as well as self-diagnosis not only based on static and perhaps configurable rules. Instead, these rules have to be adaptable by the system and according to observation of the system's states and the outcome.

Several methods from Machine Learning and Data Mining facilitate such capabilities. The analysis of huge data amounts using these methods, named Big Data Analytics, has gained a great deal of attention in the past years. The potential, not only for production scenarios, has been shown in several use cases. CMS and IIoT increase these potentials. Due to the increased data availability, these algorithms enable the system to train better models for classification, clustering, and prediction.

Methods of artificial intelligence and analytics that are suitable for CMS and IIoT as well as use cases, are discussed in:

- Chapter "Manufacturing Cyber-Physical Systems (Industrial Internet of Things)" describes the implementation of a self-learning CPS in conjunction with a knowledge database. The authors present an example that shows the planning and implementation of real physical systems using knowledge storing, complex algorithms and system structures. The described plant CPS is used for hazardous material handling, automated opening of dome covers on tank wagons for petroleum and petrochemical products.
- Chapters "Application of CPS in Machine Tools" and "Going smart—CPPS for digital production" present CPS applications for machine tools and the corresponding manufacturing processes. The former chapter includes two use cases: the intelligent chuck for a turning and the intelligent tool for milling operations. Both use cases comprise new sensor and control technologies based on analytic functions. The latter chapter focuses CPS applications for process technology on machine tools. These include, for example, the determination of process knowledge from indirect measurement signals and the corresponding visualization for the machine operator.
- Chapter "Cyber-Physical System Intelligence" focusses on systems that allow to automatically schedule, plan, reason, execute, and monitor tasks to accomplish an efficient production. Typical systems can be roughly divided in three categories: state machine based controllers, rule-based agents to more formal approaches like Golog, or planning systems with varying complexity and modeling requirements. The authors describe several approaches of all these categories and provide evaluation results from an actual implementation in a simplified Smart Factory scenario based on a group of adaptive mobile robots in simulation and real-world experiments.
- In Chap. "Big Data and Machine Learning for the Smart Factory—Solutions for Condition Monitoring, Diagnosis and Optimization" the application of Big Data platforms for factories and the modeling of formalisms to capture relevant system behavior and causalities are discussed. Further, the authors present Machine Learning algorithms to abstract system observations and give examples of the use of models for condition monitoring, predictive maintenance, and diagnosis. Finally, they demonstrate the application of models for the automatic system optimization.

- Three main milestones that have been reached in the "CPS for smart factories" activity are presented in Chap. "Overview of the CPS for Smart Factories Project: Deep Learning, Knowledge Acquisition, Anomaly Detection and Intelligent User Interfaces". First, the authors present their CPS Knowledge engineering. After that, they discuss their approach to use formal models in test scenarios to detect anomalies in physical environments. Finally, they illustrate their model based prediction with anomaly detection algorithm and the corresponding machine learning and real time verification.
- In Chap. "Applying Multi-Objective Optimization Algorithms to a Weaving Machine as Cyber-Physical Production System" the authors present a multi-objective self-optimization of weaving processes based on wireless interfaces of sensor systems and actuators. Thereby, embedded optimization algorithms enable the weaving machine to decide about optimal parameter settings autonomously. Furthermore, the weaving machine supports operators in setting up the process by providing suitable user interfaces.
- The impact of CMS and IIoT on production control and logistics is considered in Chaps. "Cyber Physical Production Control" and "A Versatile and Scalable Production Planning and Control System for Small Batch Series". The first chapter presents a general concept and first results for Cyber Physical Production Control as a means to support decision making on the basis of high-resolution real-time data. The latter chapter addresses the specific challenge of small batch sizes and presents results from the SMART FACE project from which a comprehensive CPS logistics demonstrator evolved.

5.5 Evolution of Workforce and Human-Machine-Interaction

With the introduction of CMS and IIoT the role of the today's worker will change. Competences of the future worker are focused more and more on the human cognitive capabilities. Hence, the tasks are more critical and cover for example regulating, supervising, and controlling the manufacturing process. Therefore, besides the necessity for qualification, the technical systems have to provide suitable user interfaces, enabling the user to fulfill these tasks in a proper way.

Furthermore, the interaction between human and machine advances. Collaboration between humans and machines are no more an exception. Instead, they are working as in close collaboration. These topics are covered in the following chapters:

- Chapter "CPS and the Worker: Reorientation and Requalification?" discusses the role of the future manufacturing worker. The authors demonstrate the consequences of a changing manufacturing system and give an approach how the management of a company can integrate the worker in a different way.

- In Chap. "Towards User-driven Cyber-Physical Systems—Strategies to support user intervention in provisioning of information and capabilities of cyber-physical systems" the goal is to identify challenges related to user-driven and user-defined Cyber-Physical Systems. Furthermore, the authors outline strategies to solve the identified challenges. Due to that, they describe several strategies that influence the users handling with CPS technologies.
- The technical and collaborative competency of the future employees are topic in Chap. "Competence management in the age of Cyber Physical Systems". The authors provide a categorization of different types of competency for mastering the technological and contextual complexity of CPS. In this process, a measurement instrument for these competencies is introduced.

6 A Glance into the Future: Towards Autonomous Networked Manufacturing Systems

The potentials and challenges of CPPS and CMS have already been discussed in many publications, talks, and key notes [1, 16, 26]. Nevertheless, a reference implementation has yet to be realized and several challenges still need to be solved. But, as depicted in several scenarios in this book, first steps and solutions have been realized in the past years and there are more to come.

The introduction of CMS and IIoT in the manufacturing environment will be an evolutionary process that is also triggered by innovations from other domains. In this context the book provides examples from agricultural machinery Chap. " Cyber-Physical Systems for agricultural and construction machinery—Current applications and future potential", wind energy "Application of CPS within wind energy—Current implementation and future potential", and biological tissues in Chap. "Transfer Printing for Cyber-manufacturing Systems".

In production context, the evolutionary process will sooner or later lead to networked manufacturing systems with a high degree of autonomy. Such systems provide plug and produce as well as self-optimization and self-diagnosis capabilities. They are organized in a decentralized manner, increasing robustness and adaptability. Due to a high information transparency that has to be reached in future CMS, the production will be efficient with regards to costs and resources. A flexible and adaptable production scheduling will be possible, allowing the production of very small lot sizes.

Building innovation communities that help companies and their employees to successfully go through this digital transformation will be a key factor for economic success. In this context chapter "Advanced Manufacturing Innovation Ecosystems: The Case of Massachusetts" illustrates an economic state analysis and subsequent recommendations for creating and fostering innovation ecosystems by the case Massachusetts.

Beyond these ecosystems, we need to find answers regarding societal implications as well as legal, security, and safety aspects. Furthermore, the increased dependability on technology and providers of technological solutions require established companies to rethink long grown structures.

Acknowledgments The work for this chapter and for the editing of the book has been partly funded within the Cluster of Excellence "Integrative Production Technology for High-Wage Countries". The authors would therefor like to thank the German Research Foundation DFG for the kind support.

References

1. Acatech (2011) National Academy of Science and Engineering. Cyber-physical systems: driving force for innovation in mobility, health, energy and production, München
2. Atzori L, Iera A, Morabito G (2010) The internet of things: a survey. Comput Netw 54 (15):2787–2805
3. Bledowski K (2015) The internet of things: industrie 4.0 vs. the industrial internet. https://www.mapi.net/forecasts-data/internet-things-industrie-40-vs-industrial-internet. Accessed 14 June 2016
4. CIA (2016) The world factbook: Germany. https://www.cia.gov/library/publications/the-world-factbook/geos/gm.html. Accessed 29 June 2016
5. Colombo AW, Bangemann T, Karnouskos S, Delsing J, Stluka P, Harrison R, Jammes F, Lastra JL (2014) Industrial cloud-based cyber-physical systems: the IMC-AESOP approach. Springer, Switzerland
6. CyPhERS (2013) Characteristics, capabilities, potential applications of cyber-physical systems: a preliminary analysis. http://www.cyphers.eu/sites/default/files/D2.1.pdf. Accessed 17 Mar 2016
7. Economist (2015) Germany's industry: does Deutschland do digital? http://www.economist.com/news/business/21678774-europes-biggest-economy-rightly-worried-digitisation-threat-its-industrial. Accessed 19 June 2016
8. Eigner M, Stelzer R (2013) Product lifecycle management: Ein Leitfaden für product development und life cycle management, 2, neu, bearb edn. VDI Springer, Dordrecht
9. Fortino G, Di Fatta G, Ochoa SF (2015) Future generation computer systems: special issue on "Cyber-physical systems (CPS), internet of things (IoT) and big data". http://www.journals.elsevier.com/future-generation-computer-systems/call-for-papers/special-issue-on-cyber-physical-systems-cps-internet-of-thin/. Accessed 27 June 2016
10. Gamble C, Larsen PG, Pierce K, Woodcock J (2015) Cyber-physical systems design: formal foundations, methods and integrated tool chains. In: 3rd FME workshop on formal methods in software engineering (FormaliSE), pp 40–46
11. Gartner (2015) Gartner's 2015 hype cycle for emerging technologies identifies the computing innovations that organizations should monitor. http://www.gartner.com/newsroom/id/3114217. Accessed 14 June 2016
12. GE (2016) Industrial internet insights: bring together brilliant machines, advanced analytics and people at work. https://www.ge.com/digital/industrial-internet. Accessed 14 June 2016
13. Geissbauer R, Vedso J, Schrauf S (2016) Industry 4.0: building the digital enterprise: 2016 global industry 4.0 survey. PwC, Munich
14. Gordon RJ (2016) The rise and fall of american growth: the US standard of living since the civil war. Princeton University Press, New Jersey
15. Hoang DD, Paik H, Kim C (2012) Service-oriented middleware architectures for cyber-physical systems. Int J Comput Sci Netw Secur 12(1):79–87

16. Kagermann H, Wahlster W, Helbig J (2013) Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0. Abschlussbericht des Arbeitskreises Industrie 4:5
17. Kühn W (2006) Digital factory: simulation enhancing the product and production engineering process. In: Winter simulation conference, pp 1899–1906
18. Lee EA (2008) Cyber physical systems: Design challenges. In: 11th IEEE international symposium on object oriented real-time distributed computing (ISORC), pp 363–369
19. Lee J, Bagheri B, Jin C (2016) Introduction to cyber manufacturing. Manuf Lett 8:12–15
20. Lin K, Panahi M (2010) A real-time service-oriented framework to support sustainable cyber-physical systems. In: 8th IEEE international conference on industrial informatics (INDIN), pp 15–21
21. Münchrath J (2016) Falsche Versprechen. Handelsblatt 70(110):44–45
22. NASA (2012) Cyber-physical systems modeling and analysis (CPSMA) initiative. http://www.nasa.gov/centers/ames/cct/office/studies/cyber-physical_systems.html. Accessed 24 June 2016
23. Song H, Rawat D, Jeschke S, Brecher C (eds) (2016) Cyber-physical systems: foundations, principles and applications. Academic Press, London
24. Törngren M, Asplund F, Bensalem S, McDermid J, Passerone R, Pfeifer H, Sangiovanni-Vincentelli A, Schätz B (2016) Characterization, analysis, and recommendations for exploiting the opportunities of cyber-physical systems. In: Song H, Rawat D, Jeschke S, Brecher C (eds) Cyber-physical systems: foundations, principles and applications. Academic Press, London, pp 3–14
25. VDI (2008) Richtlinie 2206—Digitale Fabrik Grundlagen. VDI-Verlag, Düsseldorf
26. VDI/VDE-Gesellschaft (2013) Cyber-Physical Systems: Chancen und Nutzen aus Sicht der Automation, Düsseldorf
27. Vogel-Heuser B, Kegel G, Bender K, Wucherer K (2009) Global information architecture for industrial automation. Automatisierungstechnische Praxis (atp) 51(1):108–115
28. Whitmore A, Agarwal A, Da Xu L (2015) The internet of things—a survey of topics and trends. Inf Syst Front 17(2):261–274. doi:10.1007/s10796-014-9489-2

An Application Map for Industrial Cyber-Physical Systems

Sascha Julian Oks, Albrecht Fritzsche and Kathrin M. Möslein

1 An Introduction to Cyber-Physical Systems

Cyber-physical systems are the foundation of many exciting visions and scenarios of the future: Self driving cars communicating with their surroundings, ambient assisted living for senior citizens who get automated assistance in case of medical emergencies and electricity generation and storage oriented at real time demand are just a few examples of the immense scope of application [11]. The mentioned examples show that cyber-physical systems are expected to have an impact in various domains such as: Mobility, healthcare, logistics, industrial production and further more. This comes along with noticeable change for citizens in their daily lives and routines on micro-, meso- as well as macro-level:

- Individuals can profit from cyber-physical systems personally (micro-level), residing in *smart homes* and supported by *ambient assisted living*. The engineering of new service systems based on cyber-physical systems, bringing together tangible and intangible resources, enable new value propositions [4].

S.J. Oks (✉) · A. Fritzsche · K.M. Möslein
Chair of Information Systems I, Innovation and Value Creation,
Friedrich-Alexander-University of Erlangen-Nuremberg,
Lange Gasse 20, 90403 Nuremberg, Germany
e-mail: sascha.oks@fau.de

A. Fritzsche
e-mail: albrecht.fritzsche@fau.de

K.M. Möslein
e-mail: kathrin.moeslein@fau.de

K.M. Möslein
Center for Leading Innovation and Cooperation (CLIC),
HHL Leipzig Graduate School of Management, Leipzig, Germany

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_2

- Users benefit from the merge of physical status information and virtual data, like in the case of traffic estimator systems processing the location and travel speed of each system participant or other *smart mobility* applications (meso-level).
- A significant expansion of industrial production, transport and supply effectiveness and efficiency completes the expected improvements (macro-level).

This is the case, not just for national economies, but for the global economy, too. The domain of the value creation based on cyber-physical systems is especially to emphasize in the industrial context; the effects on micro-, meso-, as well as macro-level exist ranging from benefits for each individual of the value creation process to entire economies. With the implementation of industrial cyber-physical systems in factories and other industrial application areas, major potentials for improvement in terms of efficiency, process organization and work design are expected. The industrial value creation is believed to proceed with a reduction of required time and costs while the quality of products and services as well as the user benefits increase [1].

This chapter wants to give orientation to practitioners and researchers about the currently visible scope of application for cyber-physical systems according to the ongoing discussion in industry and academia. It proceeds in the following way: First, it introduces technical, human and organizational dimensions of industrial cyber-physical systems. Second, it describes categories with high potential for improvement in industrial practice by the introduction of cyber-physical systems. Third and finally, it links these categories to specific spheres and consisting application fields within the industrial value creation process. The findings are displayed in an application map, which illustrates the overall connectedness and interrelation of the spheres smart factory, industrial smart data, industrial smart services, smart products, product-related smart data, and product-related smart services and the particular application fields therein. This application map offers decision makers a compendium of application fields for industrial cyber-physical systems, which they can use as a template for their own business situation.

2 Foundations of Industrial Cyber-Physical Systems

Lee [16] lays the groundwork for the technical understanding of cyber-physical systems by describing them as “integrations of computation and physical processes”. Their application in practice, however, does not only have a technical dimension, but also a human dimension with respect to the people who use them, and an organizational dimension with respect to the surrounding economic structure. The following section gives a brief overview of the foundations of industrial cyber-physical systems in these dimensions.

2.1 Technical Dimension of Cyber-Physical Systems

From a technical point of view, cyber-physical systems are built upon the modular logic of embedded systems. Embedded systems are information processing devices which form often miniaturized components of larger computer systems. Every component has a specific functional purpose. In combination with each other, they determine the value proposition of the entire system [22]. Popular examples of embedded systems include cars, household appliances, entertainment electronics and many more. Before the times of ubiquitous computing, they were self-contained devices with limited sensor technology and marginal interconnectedness. Comprehensive intersystem organization and linkage based on context-awareness and adaptiveness leading to self-configuration, ambient intelligence and proactive behavior was missing. These characteristics became reality with *smart objects*, entities that have a definite identity, sensing capabilities of physical conditions, mechanisms for actuation, data processing ability and networking interfaces [10]. In order to equip embedded systems with digital intelligence to extend their dedicated functionality and thus, to make them parts of cyber-physical systems as beforehand described smart objects, certain extensions are necessary.

The first requirement is the installation of sensors, which allow the digitization of physical conditions. Sensors are available for a broad range of physical phenomena. The wealth of information collected about the physical environmental conditions can be as simple as the pure occurrence detection extending to the measurement of detailed values and grades about the phenomena. Each sensor should be chosen depending on the aspired exactitude of the state description based on task and the usage context of the to equip object. The ongoing miniaturization of the previously described technical components continuously extends their scope of application. The data aggregated by these sensors needs to be processed by the local processing capacity of the smart object. Decentralized computing entails an increase in the pace of data processing while simultaneously reducing data throughput within the network infrastructure. Subsequent centralized data evaluation, in form of big data processing enables the use of the gathered data for pattern recognition and forecasts based on the recognized patterns. Hence, in cyber-physical systems, decentralized real time computing of operative measures complements centralized data evaluation for developing strategic measures.

Furthermore, communication interfaces are necessary to merge self-contained embedded systems to cyber-physical systems. In addition to previously established interfaces like Ethernet and Wi-Fi the extensive implementation of RFID, GPS and near field communication technologies allow the interconnection of a myriad of objects [26]. In parallel to this development, the introduction of the internet protocol version 6 (IPv6) solves the obstacle of an insufficient global communications network. With this new protocol, the hypothetical interconnection of approximately 340 sextillion objects via the internet is possible [17]. The upgrade of industrial machines with machine communication protocols like the OPC Unified Architecture (OPC UA) ensures the interoperability of machines from various manufacturers [19].

Nonetheless, there are several interpretations of cyber-physical systems especially when it comes to visions for their utilization in different domains. In this context, the following agendas and roadmaps are good examples of the possible variety in domains: *Living in a networked world—Integrated research agenda Cyber-Physical Systems* (*agendaCPS*) from acatech [11], *CyPhERS—Cyber-Physical European Roadmap and Strategy* funded by the EU [8], and *Strategic Vision and Business Drivers for 21st Century Cyber-Physical Systems* from the National Institute of Standards and Technology of the US Department of Commerce [24].

2.2 Human Dimension of Cyber-Physical Systems

Besides the aforementioned technological preconditions, there is a human dimension to consider. The success of the introduction of cyber-physical systems depends significantly on the acceptance by the users. The interaction between people and cyber-physical systems differs depending on the type and function of the regarded systems.

Interactions with a high level of attention and awareness are performed with the use of human-machine interfaces. These human-machine interfaces have many different forms, naming classic computer input or voice control as examples. Especially the usage of mobile devices like smartphones and tablets as control devices offers great potential for the interaction between users and cyber-physical systems. The interactions have two aspects: First, smartphones and tablets have become a commodity in many societies due to a high value in use. The wider diffusion rates of smartphones compared to desktop PCs emphasizes this trend [12]. Second, mobile internet connection allows system usage without being tied to a specific geographic location. Moreover, with operating systems that allow the installation of third-party apps, smartphones and tablets are the ideal platform technology. In many cases of human-machine interface design, the focus is thus set on the software, since the hardware is already available in the form of mobile devices.

Mobile devices, especially *wearables* (wearable technology) also contribute to passive or unconscious interactions with cyber-physical systems. As mentioned previously, it is not just smartphones and tablets, but also smart watches and fitness trackers that have become widely accepted companions of users in daily life. Moreover, virtual reality interfaces are also increasing utilized. With ubiquitous computing and emerging smart environments, carried devices communicate with the overall system in the background unnoticed by the user. By sending parameters like location, travel speed, and destination, services such as traffic based navigation or smart home systems adapt corresponding to the user's preset preferences (e.g. travel route, room temperature, etc.). In professional surroundings, the same technology can be used for safety monitoring. Usage scenarios for this are construction and maintenance activities in industrial settings. Whenever personnel working in hazardous environments remains in a position unchanged for too long, the system alerts a rescue crew automatically. Of course, compliance with data security and protection of privacy need to be a matter of fact regarding this topic.

Besides the wide distribution of mobile devices, it is the people's familiarity with using such technology in both private and professional contexts which leads to the expectation of high acceptance and adoption rates for them as human-machine interfaces for cyber-physical systems [7]. In sum, both the technical architecture and the user integration appear to provide a solid basis for the development and the implementation of cyber-physical systems in the contemporary scenario. Based on previous achievements and the ongoing advances, the remaining challenges seem manageable.

2.3 Organizational Dimension of Cyber-Physical Systems

Technically driven approaches tend to neglect that the organizational dimension plays an important role for the application of cyber-physical systems as well, particularly in a professional context. The introduction of cyber-physical systems poses as a challenge for many companies because they have to consider several layers (technology, divisional structure, business model, etc.) of the enterprise architecture at the same time. The organizational dimension with need for consideration in this process is described as follows.

Only in rare cases, cutting-edge technologies are introduced by building new production facilities solely designed to reach the maximum potential of the innovations. What usually happens is that the new technologies are integrated into an existing operational environment and thus they have to be aligned with other infrastructure [32]. For this purpose, machines need to be updated, and digital communication standards which allow the orchestration of new and old hardware at the same time need to be established. The changes in the production processes are most likely to have further effects on the structures of managerial processes and subsequently the organizational structure, as working times, supply, control routines etc. have to be revised. Therefore an effective change management does not only need to consider engineering but also business adjustments in the course of the introduction of cyber-physical systems. Companies are well-advised not to perform these adjustments in a reactive mode with respect to their overall strategy, but proactively in order to make use of the full potential offered by cyber-physical systems. New production processes and the opportunity to expand the existing range of products with new *smart products* allows extensive enhancement of the existing business model.

Especially hybrid and interactive value creation offer great potential, in this context. Hybrid value creation describes the combination of physical products with data driven services to service bundles [34]. Due to continuous points of contact between company and customer and a serial payment model, this approach is a key to long lasting customer ties accompanied by long-term income streams. Interactive value creation defines the process of cooperative collaboration between manufacturers and their customers in order to achieve a more user-oriented approach of value creation ultimately leading to products and services with a higher benefit for customers [29]. The simultaneous practice of both approaches offers increased usefulness actuated by the mutual enhancing effects of either procedure. Despite such advantages, many companies perceive these modifications to the established and existing business models more as a challenge than a chance. In case of cyber-physical systems, this is intensified by the potential changes that have to be considered simultaneously in the manufacturing process, the product portfolio and new, to be thought of, services all at once.

Nevertheless, despite the previously mentioned advantage of the common use of mobile communication devices for the user integration for cyber-physical systems, the introduction of new technologies and procedures generates a need for adjustment for the personnel. In most cases, these adjustments consist of changes in work routines and procedures that entail training courses and other qualification measures for the workforce. Since personnel might perceive such activities as additional efforts besides the usual tasks, it is a managerial challenge to clarify the resulting benefits for personnel and to motivate them to adopt the new technologies.

The availability of far more data than before, due to cyber-physical systems and smart products, offers companies a variety of exploitation scenarios (predictive maintenance, hybrid value creation, big data solutions, etc.). In this context, the potential of inter-organizational data exchange is to highlight for integrated logistics concepts, just-in-time production etc., as it brings a new efficiency level to inter-professional production networks. This can foster strategic alliances in between corporations while at the time reducing lock-in effects and stimulating markets. Inter-company cooperation on this level requires a major exchange of data in real time. For many companies, this seems synonymous to an inestimable risk of data loss, offering a wide range of potential targets for hacking and industry espionage [31]. The step toward larger collaboration across company boundaries is therefore often difficult to take for them. However, there are advanced cybersecurity standards available that can help prevent hacking and espionage effectively, if they are combined with a suitable data sharing strategy by the company [28]. This illustrates one more time the importance of the organizational dimension of cyber-physical systems.

Like any other rollout in the industrial context, the introduction of cyber-physical systems means an investment of financial capital. Based on the multitude of factors to consider, it is nearly impossible to take all into account at the same time without a systematic approach. The identification of proper application fields matching the unique and specific needs of an organization and the estimation of the overall benefit is difficult and it is even harder to estimate reliable figures of the return on investment. While this is already deterrent for large-scale enterprises, it especially hinders SMEs to utilize cyber-physical systems and the optimizations associated therewith [30].

In comparison, the technical and human dimensions of cyber-physical-systems seem more advanced, while the organizational dimension is still in lack of maturity. Figure 1 gives an overview of the current scenario and its various aspects.

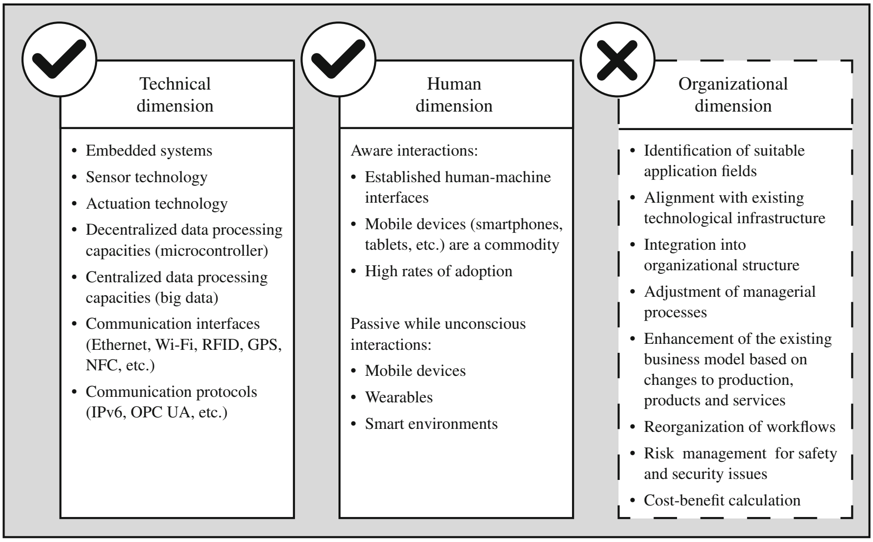


Fig. 1 Dimensions of the successful implementation of industrial cyber-physical systems

3 Categories of Potential Improvement for Industrial Cyber-Physical Systems

Like mentioned before, the adoption of cyber-physical systems in the industrial context offers great potentials ranging from benefits for each individual of the value creation process to entire economies. The expected effects include sustainable growth of a nation's GDP, accompanied by an increase of individual wealth and living standards [11]. Governmental institutions in many countries have recognized these highly promising anticipations and have therefore implemented funding initiatives with the goal to stimulate the adoption of cyber-physical systems in the industrial sector of their countries. While most of the worldwide public initiatives pursue this general aim, they differentiate in design and implementation structure as well as funding volume. Prominent examples are the following:

- In the United States, the National Network for Manufacturing pursues its initiative *Advanced Manufacturing Partnership 2.0* with the objective to "use new, often leading-edge machines and processes to make products that are unique, better, or even cheaper. Advanced manufacturing also facilitates rapid integration of process improvements, readily permits changes in design, such as new part features or substitute materials, and accommodates customization and cost-effective low-volume production." [20].
- The German initiative *Industrie 4.0* aims to strengthen the position of its mechanical engineering sector as a global market leader. Moreover, there is a focus on developing norms and standards for communication protocols as well

as providing SME-specific guidelines for the implementation of innovative technologies. The “4.0” symbolizes the great expectations attributed to the new technologies, lifting this development on a level with the three former industrial revolutions [6].

- *Catapult—High Value Manufacturing*, the initiative in the United Kingdom, strives among other things, to foster further digitization in manufacturing processes and to reinvigorate the industrial production that has been declining in the UK over the last decades [14].
- China aspires to update its manufacturing industry with the program *Made in China 2025* to leave the times of being the “workbench of the world” behind. The main goal is a better overall innovativeness in combination with superior quality of the manufactured products. Moreover, there is a focus on a more ecologically reasonable economic progress and education of native specialists [18].

Besides governmental initiatives, there is a multitude of initiatives and platforms run and funded by the private sector. To emphasize are e.g. the US based *Industrial Internet Consortium* or the *Industrial Value Chain Initiative* from Japan [21].

The frequent referral to cyber-physical systems as a key component of the implementation of *smart factories* in initiatives of both developed as well as emerging economies underlines once more the importance of these technologies.

Despite all the public attention and financial support, it remains widely unclear for many decision makers how cyber-physical systems can actually generate benefit for their companies in practice. The findings of this book chapter were achieved within the research project “Resource-Cockpit for Socio-Cyber-Physical-Systems” funded by the German Federal Ministry of Education and Research. In the course of the project experts out of the fields of management, industrial associations, research, labor unions and work committees as well as the federal employment agency were interviewed following a qualitative research design. The perception of the topic by the shop floor personnel was included via focus groups. The analysis of the interviews and focus groups in combination with desk research led to the upcoming categories and built the foundation for the elaboration of the application map. Before describing the actual fields of application for cyber-physical systems, the main categories in which the experts foresee high potential for improvement in industrial practice are listed.

These categories are automatization, autonomization, human-machine interaction, decentralization, digitization for process alignment, big data, cyber security, knowledge management and qualification. An overview of these categories of potential improvement is given in Table 1.

Table 1 Categories of potential improvement for industrial cyber-physical systems

Automatization	• Integrated flow of production • Machine-to-machine communication (M2M communication) • Plug-and-produce machinery interconnections • Automated guided vehicles (AGV)
Autonomization	• Supervisory control and data acquisition (SCADA) • Condition monitoring • System reconfiguration
Human-machine interaction	• Unrestrained human-machine collaboration • Robotic exoskeletons • Decision support systems • Resource cockpits • Augmented reality
Decentralization	• Decentralized computing in modular networks • Complex event processing
Digitization for process alignment	• Digitization of warehousing and logistics • Automated e-procurement • Industrial services in the field of maintenance, repair and operations (MRO) • Digital image of products • Document digitization
Big data	• Pattern detection • Data processing warehousing solutions
Cyber security	• Cyber security solutions • Engineering of safety system infrastructures
Knowledge management	• Systematic recording, categorizing and mapping of implicit knowledge • Action guidelines
Qualification	• Qualification concepts • E-learning

3.1 Automatization

Industries in developed and emerging countries rely strongly on a highly developed manufacturing process as a basis for their success on the market. This includes the extensive usage of technology in various ways and its automated operation. Over time, the motives for automatization have changed: Coming from the goal to lighten the workload of employees, automatization soon raised productivity due to new procedures of product assembling. Taylorism in 19th century and computer integrated manufacturing (CIM) in the 20th century are the most prominent development periods in the past [13]. Cyber-physical systems offer the potential for the next large developmental step in the application of automatization. Based on the stated foundations the following configurations can be identified.

The *integrated flow of production* profits from the situational context awareness of smart machines and smart production materials. A digital image of the product to be assembled is stored on a miniaturized data carrier attached to each production material. Whenever a production step is about to be executed the machine reads the production instructions out of the data carrier and processes the production material as required. In this way, automated batch size one production becomes executable. To prevent uneconomic sizes of lots and suboptimal retooling cycles, *machine-to-machine communication* (*M2M communication*) has an elevated importance in this context. Machines within one line of production exchange information about pending steps of procedure and optimize the sequence holistically based on previous determined algorithms. Moreover, the inter machinery communication is a key for the establishment of *plug-and-produce machinery interconnections*. Based on task and order, different compilations of machinery are needed to fulfill required process steps. In static assembly lines, this can mean that certain machines are unused but still not available for task performance. In the case of plug-and-produce machinery interconnection, only the needed machines are compiled to an assembly line. Vacant machines can be used for other tasks synchronously while machines with a malfunction can be exchanged easily. The basic requirement for these constantly changing machinery networks are cross vendor communication standards.

Besides the use in the production process itself, cyber-physical systems offer improvement potential for production supporting processes. *Automated guided vehicles* (*AGV*) interact via sensors and actuators with their environment and fulfill tasks like the transport of component groups and working materials as well as warehousing. The full potential benefits of automated guided vehicles become available when they are integrated into the network of the before mentioned machine-to-machine communication.

3.2 Autonomization

The term autonomization closely relates to automatization but is not an equivalent. Autonomization stands for the approach to control and coordinate automated processes without external (human) intervention but by system internal evaluation mechanisms. Based on self-optimizing algorithms the overall production system anticipates critical incidents and other occurrences of the operating history and optimizes the solution behavior.

A new level of *supervisory control and data acquisition* (*SCADA*) becomes possible in this way. Opposite to the up to now approach, based on continuously available real time data the future SCADA allows detailed *condition monitoring* and situation based *system reconfiguration*. Automated debugging in case of severe failure conditions is another advancement in this context. This offers both economic likewise work safety improvements: The automated batch size management facilitates the cost-efficient production of mass-customized products based on individual customer needs. In addition, autonomic procedures allow the reduction of the number of human operators. Besides the reduction of labor costs which enables competitive production in high wage economies [5], autonomic production can be a partial solution for the demographic change in western societies due to a declining workforce [15]. Furthermore, autonomization has obvious implications for the safety in manufacturing processes, especially during dangerous stages of fabrication, and while processing components and materials containing hazardous substances. The absence of personnel eliminates the danger of work accidents.

3.3 Human-Machine Interaction

Although the previous section noted certain advantages in the reduction of deployment of staff, the introduction of cyber-physical systems will not make men replaceable in the industrial context. Humans are still superior to machines in certain tasks and activities. In other areas, while machines could replace workers it would mean a financial disadvantage. Therefore, user integration is essential for the successful implementation of cyber-physical systems wherever humans are part the system or interact with it.

While today in general, due to safety regulations, machines and humans work physically separated from each other, cyber-physical systems allow unrestricted interaction. Sensor equipped shells which overlay machinery parts register contact in between machines and workers within milliseconds and stop harmful movements. Camera systems tracking positions and movements of both workers and machines are another method to prevent collisions and make protecting fences obsolete. The *unrestrained human-machine collaboration* enables each party to unfold their inherent strengths leading to an overall optimization. In addition, there is a high potential to reduce the workload for the personnel wherever physical strength is needed. Enabled by wearable support systems like *robotic exoskeletons* lifting and carrying activities become less tiring for the body [3]. Robotic exoskeletons have a positive influence on both the performance and the overall working lives of the personnel.

Besides the mentioned direct cooperation and interaction between humans and machines in fulfilling physical tasks, cyber-physical systems can be the basis for service systems [4]. In form of *decision support systems* users are supplied with needed data and information relevant to execute their job. When engineering these service systems several matters need to be considered: First, industrial processes include personnel with different positions of the organizational hierarchy with distinguished tasks. Therefore, a comprehensive role model should be utilized when conceptualizing these decision support systems [25]. By doing so, every role gets the appropriate reading-, writing-, or administrator rights. This ensures the supply with required information for task fulfillment while protecting the system from unintentional entry and outflow of critical information. Second, the right amount of information offered needs to be determined. Due to the plentitude of data gathered by machine attached sensors and other sources, an unfiltered supply of this data easily leads to an information overflow. Therefor the decision support systems need to be built on an evaluated reference architecture bringing together the knowledge about each task and the valuable information for the fulfillment of it. By doing so, each role is supplied with right data based information at the right time without the need for across systems information procurement. Third, the in this way composed *resource cockpits* do not just need a useful concept of information supply but also the right hardware and visualization methods for an optimal utilization. To integrate the information supply into the work flow ideally, solutions like *augmented reality* are very suitable. Using hardware like data glasses or other wearable technology, information can be presented as a graphic overlap over physical infrastructure. Operating data, work instructions and error localizations can be presented virtually in semantic context with real-world artefacts like machines. An additional advantage is that augmented reality enables the presentation of before mentioned information about covered modules of machines by offering a virtual insight into the machine without opening it physically.

3.4 Decentralization

Production aligned with CIM-standards is mainly based on centralized hierarchically structured computing processes. This owes to the characteristic of the hardware and software which was standard as the concept CIM was developed as well as to the at this time prevailing business logic. While in certain scenarios centralized data processing is still advantageous (e.g. big data analytics), for real time relevant tasks like production execution, *decentralized computing in modular networks* offers unequivocal benefits. Cyber-physical systems strongly link to the approach of decentralization [33]. Enabling factor is the continuous miniaturization of technical components along with an increase in processing power of these. Thus, c*omplex event processing* is no longer bound to centralized computing units but can be performed in a leaner and faster way based on decentralized computing network solutions.

3.5 Digitization for Process Alignment

The sensor based digitization of physical operational sequences of the production process only unfolds its full potential as part of an entirely digitized factory. Hence, digitization should be fostered not just in the core production process but also in all production supporting sectors. An extensive *digitization of warehousing and logistics* based on RFID or NFC systems enables self-organizing production networks to include the real time inventory into the production program. Furthermore, continuous inventory and stocktaking based on sensor and actuator equipped pallets, boxes, shelves and also production materials enable an *automated e-procurement*. This improves the in time availability of parts and materials delivered by suppliers and allows an extended use of just-in-time production. In this way, strategic partners like subcontractors can be integrated more profoundly into the value chain. This is not just the case for suppliers but especially for providers of *industrial services in the field of maintenance, repair and operations* (*MRO*). With comprehensive information and the option of remote control, several MRO activities in case of software or operational errors can be conducted from a distance. In the event of physical defects the maintenance personnel of the machine operator can be supported by experts of the machine manufacturer's company who can base their advice on real time data received via a secure connection.

On basis of the fact that in most cases the introduction of cyber-physical systems does not proceed in form of the construction of completely new production facilities from scratch but as a transition with an update of the existing machine fleet, digitization has to be considered as well from this point of view. When introducing new decision support systems to the plant personnel the requirement has to be that all relevant information can be made available via one decision support system on a single device. Media discontinuities are perceived as cumbersome by the user. Additional time consuming research work for e.g. blue prints or handbooks in paper-based filing systems and archives lead to only modest adoption rates of the decision support system. To counteract this, all relevant documents like handbooks, blueprints, protocols, etc. should be digitized. The act of *document digitization* needs to be completed by inventorying the content of the files to make the option of searching the document available.

Besides the advantages for the organization of the production process, digitization offers applicability for product improvements as well. The before introduced *digital image of each product* stored on a microchip attached to the product which is used during the production process for the communication between the to be assembled product and the executing machines, maps afterwards the individual product life cycle. With data of usage and every after-sales service, it provides valuable insights which can be utilized in form of further product development and offerings of product-enhancing services.

3.6 Big Data

The extensive installation of sensors on machines causes a massive increase in the volume of data collected within industrial processes. The data consist of operating data, error lists, history of maintenance activities and alike. In combination with the related business data, the overall plethora of data provides the raw material for process optimizations and other applications. To set this potential for optimizations free, the raw data needs to be processed systematically, passing through various algorithms. The results are prepared information with specific application objectives. Especially *pattern detection* is to mention in this context, since this method identifies and quantifies cause and effect correlations and allows predictions of state changes. The significance of the information given out by the analysis depends on the amount of data processed. Therefore, it can be in the interest of individual companies to unite their data sets with the goal of a joint asset in form of more precise and meaningful informational results. Requirements for these joint operations are *data processing warehousing* solutions with extremely large computing and storage capacities.

3.7 Cyber Security

Many of the before described categories of potential improvement have in common that the functioning of the introduced cyber-physical systems is dependent on data interchange between separate system components. In various cases, the data interchange does not just proceed within enclosed IT systems but also web based across company boundaries. Especially in case of close integration into value-adding networks or in interdependent systems of systems, a widely distributed data flow is a fundamental prerequisite. The extended value in use comes with the risk of an increased vulnerability due to cyber threats. These cyber threats consist of data theft, sabotage, industrial espionage, and further more. In the event of a successful outcome of these digital attacks, the negative consequences for the affected companies are incalculable. The range includes malfunctioning machines, an endangering of the work safety up to the loss of customer confidence. These alarming consequences underline the need for reasonable *cyber security solutions*. A reliable security concept should consist of measures both on individual system participant's level as well as on the overall system's level [9].

Especially for the scope of direct cooperation and interaction between personnel and machines, manipulability needs to be eliminated. Therefore, the *engineering of safety system infrastructures* is a notable aspect with regard of operating cyber-physical systems.

3.8 Knowledge Management

Among other things, cyber-physical systems enable an increased level of effectivity and efficiency in the industrial value creation because of the amount of real time information they provide about technical processes. However, for the utilization of the full potential of cyber-physical systems the collected information should include non-technical sources of data, too. Implicit knowledge of the personnel falls into this category. Activities proceeded during the work process are based on the practical knowledge of the staff. In many cases this knowledge is only available informally and difficult to be formalized. However, due to the great value of this knowledge, methods should be introduced to systematically record, categorize, and map it. The availability of this knowledge can be used for the design of *action guidelines,* which are an essential part of decision support systems. An example can be the repair of a malfunctioning machine. When an error occurs for the first time, the problem-solving process should be documented, so when it occurs for the next time an action guideline is available and whoever fulfills the repair, can profit from the experience curve effect. However, the process of *systematically recording, categorizing, and mapping implicit knowledge* implies an additional effort for the employees. Consequently, it is essential to clarify the overall added value based on the availability of the action guidelines once the decision support system is engineered and implemented. Incentive systems are a proper instrument to ensure the participation of all involved stakeholders.

3.9 Qualification

The implementation of cyber-physical systems entails major change in the process of industrial value creation. This affects in many areas the role of men within this process as well. Tasks, roles, and requirements of the personnel pass through a major transformation. Education concepts and study contents of apprentices need to be adjusted to the new needs. A particular challenge in this regard is the retraining and teaching of content to the existing workforce. Methods for employee motivation and integration into new training measures are necessary. Sometimes even longstanding customs, biases, and other means of resistance need to be managed. The elaboration of new *qualification concepts* for both trainees as well as experienced staff, ensuring the ability to operate and interact with cyber-physical systems, are an important measure for a successful change management.

Beyond the recording, categorizing and mapping of implicit knowledge and the digitization of information that was formerly decentralized and difficult to access, it enables the introduction of new *e-learning* methods. E-learning offers are exploited by the use of mobile devices as human-machine interfaces, since they can also be used for this purpose.

4 Elaboration of an Application Map for Industrial Cyber-Physical Systems

In this final part of the chapter, the pointed out categories with high potential of improvement are matched with specific spheres and inherent application fields of the industrial value creation process. To structure these application fields the following spheres categorize them: Smart factory, industrial smart data, industrial smart services, smart products, product-related smart data, and product-related smart services. Even though the spheres of smart products as well as product-related smart data and product-related smart services do not directly belong in the industrial sector, they have strong interdependencies with it and influence the introduction of cyber-physical systems in industrial processes significantly. Therefore, the application fields that fall into these spheres will be illustrated as well. Foregone, the strong interconnectedness and combined effects of the spheres and application fields are to emphasize. Only a few applications fields within these spheres can be classified as stand-alone application scenarios.

As the core of cyber-physical system based industrial manufacturing, the sphere smart factory will be approached first.

4.1 Smart Factory

The fabrication and assembling of products and the underlying and contributing processes in the smart factory offer a great variety of application fields for cyber-physical systems. First to mention is the *production* itself. Production planning and control have to take more factors into account than before and orchestrate a great amount of technical, mechanical and digital processes with minimal tolerance of process time. Therefore, the production management needs to achieve a new level of automatization and autonomization. To reach the requirements of a forward-pointing and competitive production planning and control, these systems should be self-(re)configuring, self-optimizing, adaptive, context-aware, and real-time capable. To reach this overall goal, cyber-physical systems should be installed throughout the *assembly line*. In particular, the implementation of features in the area of automatization and autonomization (machine-to-machine communication, plug-and-produce machinery interconnections and automated guided vehicles as well as supervisory control and data acquisition and system reconfiguration mechanisms) are promising. Moreover, the assembly line is the application field for most cyber-physical systems allowing an integrated human-machine interaction. Jointly these measures lead to a reengineered production procedure allowing the economic manufacturing of batch size one.

To ensure an integrated flow of production, further application fields offer great potential for the implementation of cyber-physical systems. Incoming *logistics* are one of these. An automated e-procurement ensures a sufficient inflow of production materials and precursors. The optimum of order quantity is automatically calculated with real-time data from production, warehousing, and incoming orders. Moreover, market trends, price developments and other company external data can be integrated for an optimized e-procurement. With strategic suppliers and subcontractors, an integrated supply chain can be established based on cyber-physical systems. For this purpose, the interwoven production processes of several companies can be linked virtually to a strategic production network.

Once the production materials and precursors arrive at the smart factory, a cyber-physical system based *resource management* ensures the automated influx of these into the production process. Automated guided vehicles collect the means of production from warehouses with virtual commissioning. Another field of application in the context of resource management is the alignment of production with smart grids. In these intelligent electricity networks, the production of energy is closely tied to the actual demand [2]. Depending on current outstanding orders and potential future orders, a cyber-physical system based energy management can schedule energy intensive stages of production for timeframes with favorable electricity rates. The general increase of efficiency both in processes and resource usage combined with the optimized energy consumption allows cost reductions and a more "green production" at the same time.

The *quality management* profits of the use of cyber-physical systems, too. With real time data from the production process as well as from products in use (especially of smart products), deviations to estimated values throughout the production process can be detected precisely. This contributes to the continuous quality assurance of the production but also supports the understanding of causes of product failure and linking it to manufacturing problems.

Research and development profit in an analogical manner of the wide spectrum of data availability due the application of cyber-physical systems within production and smart products. A digital image of each product stored on a microchip attached to the product, holding record about assembling, services activities, repairs and other related incidents of the individual product lifecycle, allow an evaluation of product's strengths and weaknesses. These conclusions are helpful for the continued development of new product versions. Furthermore, data from products in use is valuable for this purpose. The ways and manners how customers use the products, give an overview how well the product is aligned to customer needs.

The application of cyber-physical systems is also beneficial for the customer relationship management: In the context of *distribution*, the customer can keep track of his order until it arrives. While for standardized products this is nothing new, for individualized and custom-made products an extension can be made to the present shipping tracking. For customized products a tracking through the entire manufacturing process becomes available due to the application of cyber-physical system along the assembly line. Since the traceability of every order is a requirement for the automated production procedures, it can be converted to a service for the customer as well. By doing so, the customer cannot just track the order through the production but can also still modify it during the production for forthcoming production steps. The *value proposition* can be extended to further areas. The new manufacturing capabilities due to the application of cyber-physical systems enable the extensive production of smart products with potential for an extended customer benefit. The specifics of smart products and the interconnection of them to the smart factory will be described in the upcoming Sect. 4.4 (smart products).

Before that, the focus is directed towards *industrial smart data* and the generation of it.

4.2 Industrial Smart Data

In the previous section, application fields for cyber-physical systems in the smart factory were described. Remarkable is the broad variety but also the indirectly affected business units profiting from the application. What all application fields have in common is the generation of large amounts of data. However, all the accruing data captured by sensors installed in the smart factory is only then from value, when it is stored, processed and aggregated and thus transferred into contextualize information.

For the reason alone of the sheer number of sensors and the amount of data collected by them in the smart factory, special *industrial data warehousing* solutions are in need. Therefore, when companies apply cyber-physical systems within their production and surrounding application fields, a connected adequate data storing solution is essential.

The continuously inflowing and then stored data needs to be processed and interpreted. Like described in the section about the smart factory, there are several contexts for which the analyzed data can be utilized. To achieve this objective in a systematic way, the application field of *process engineering for industrial data analysis* is appointed. The continuous development and advancement of algorithms to process the data to valuable information is the main task of this application field.

The elaborated algorithms are employed in the process of *industrial data analysis*. In this application field, the data sets from different sources within the smart factory are evaluated and interpreted. The focus of these actions is the detection of data patterns which can be correlated to certain events. The determination of the likelihood of occurrence and the deduction of forecasts is a further ambition of these activities. Overall, the process of industrial data analysis can be summarized by the term “big data to smart data”.

In certain cases, the information resulting from industrial data analysis is not meaningful enough on its own. In these cases, the required information cannot be extracted exclusively out of the data pool generated by the factory internal cyber-physical systems. In order to fill this gap, *industrial data enrichment* needs to be applied. The concept of industrial data enrichment can be described as followed: Depending on the task to be fulfilled and the availability of data within the company, external data sources are identified and added to the database. Examples of these external data are market analysis, economic and political forecast, exchange rates or alike. Moreover, collected data from the manufactured products that are now in use are to mention in this context. The used data sources can be both free of charge or payable services.

Another case of missing data can be attributed to the reason that certain data exists within the company but not in a suitable form. This is the case, if documents are only available as hard copies or processing steps are executed with media discontinuity, leaving data in an analog form. To address this problem, methods for systematic *digitization* are necessary. However, the process of digitization goes beyond the pure activity of transferring information from an analog in a digital state: The systematic tagging and filing of the new digitized data ensures the finding and utilizing of it in a practicable way.

The applications of cyber-physical systems create and require great amounts of data at the same time. To ensure unhindered process sequence and flow of data, the interconnection of all involved cyber-physical systems is required. In certain scenarios like strategic production networks, this means an exchange of information in between independent companies via the internet. To secure safety and security, *industrial cyber security* is an application field to emphasize for the safe operation of cyber-physical systems.

4.3 Industrial Smart Services

The information and conclusions gained from industrial smart data do not only directly reenter the production execution process but constitute the foundations for a broad range of *industrial smart services*, too. These data driven services can be in-house services, supporting the own value creation processes or services offered to external customers. Therefore, the gathered data can be seen as an enabling foundation for new services, which have the aim to further optimize the value creating process. Besides the smart data based services, there are services, e.g. qualification courses, which operate with limited usage of data. Both smart data intensive as well as less data requiring, internal and external service offerings are described in the following precisely.

The application of industrial cyber-physical systems is often associated with the opportunity for the enhancement of existing business models or the creation of entirely new ones. Therefore, the conclusions gained out of the industrial smart data can be used for *business model development*. The availability of detailed information about both production processes and products in use enriched with data from other contexts, facilitate the systematic development or adjustment of business models.

While the application field of business model development shows the potential for strategic usage of smart services, there are also operative scenarios. In this sense, *employee qualification* is a necessary action to enable a functioning integration of users into cyber-physical systems. The compiling and execution of contemporary training concepts ensure the familiarity and appropriate interaction of employees with cyber-physical systems.

Based on conducted employee qualification measures and systematic user integration into cyber-physical systems, advanced forms of *knowledge management* can be introduced. The objective of these knowledge management systems is to gather implicit knowledge of employees for a reintroduction in case of need. By doing so, the implicit knowledge of the staff becomes another data source for the application field of industrial data enrichment. To assure the willingness of the workforce to contribute to these knowledge management systems the process of knowledge collection must not be unnecessary disruptive and the benefits offered must outdo the effort.

A very illustrative example for the advantageous utilization of knowledge management systems is *maintenance*. Maintenance activities aim to assure the availability of production capacities. They include upkeeps and inspections during the running process as well as repairs and overhauls in the case of malfunctions and errors. While the handling of recurring task in the field of upkeeps and inspections are standardized and scheduled, the repair of malfunctions and the solving of errors can be considered as a predominantly diverse with a high degree of freedom in execution. Especially when malfunctions and errors with a high complexity come in presence, knowledge from previous occurrences about the solving comes in hand. Ideally, this knowledge is presented in a structured way in form of an action guideline. Resource cockpits are a suitable platform for the accumulation of these and other context based information provided to the maintenance personnel. The value in use of the resource cockpit increases over time since every solution to a malfunction or error is entered into the system and linked to an event (collection of industrial smart data). Whenever the malfunction or error occurs again, the assigned worker can profit of the preparatory work of colleagues. Overtime the positive effects of a non-personal learning curve set in.

Besides the described potentials for maintenance due to advanced knowledge management, cyber-physical systems can be applied to improve the overall maintenance process. The objective is the reduction of machine downtime by continuously analyzing the condition of the machinery components (condition based maintenance). Entering both the data collected by the installed sensors of all machinery components and the occurrences of errors into the industrial data analysis, patterns, and causal correlations can be identified. Based on this information the accuracy of predictive maintenance can be improved. The application of predictive maintenance can have a positive effect on the availability of production capacities due to fewer disorders in the production process and optimized periods of use of each machinery component. Furthermore, the application of cyber-physical systems enhances the use of remote maintenance. Based on the vast availability of information extracted from industrial smart data, remote activities to solve problems or to support personnel which are at the scene from a distance can be offered.

All previously introduced industrial smart services can be implemented as in-house solutions but also as services offered to external companies as service seekers. The market commercialization of *industrial service systems* provides an opportunity to gain further financial returns based on cyber-physical systems. These services range from consulting activities to strategic cooperation between manufacturer and service provider within production or data evaluation.

4.4 Smart Products

Besides the until here presented potentials within the several application fields of the so far introduced spheres, the industrial value creation can profit significantly due to the integration of cyber-physical systems into the after sales period of the product life cycle.

Accordingly, smart products and the related smart data and smart services in the customer context offer the possibility to maintain a continuous connection between the customer and the product in use on the one side and the manufacturer on the other side. The benefits of this after sales connection accrue for both the manufacturer and the customer. The manufacturer receives information about how customers use their products and can therefore align future hard- and software design due to customer needs and give out updates if necessary but most importantly adjust the production process if malfunctioning of products in use is detected. The product quality is hereby improved continuously. So the business units of marketing, product development, and production benefit from the described data backflow in general. Of course anonymization and data security are the fundamental prerequisite for these procedures. The customer also profits from the cyber-physical components of the product. This becomes clear when analyzing the characteristics and *functionalities* of smart products. With identifiable, situated, pro-active, adaptive, context-aware and real-time capable the attributes of smart products are very similar to those of the production mechanisms in the smart factory. Based on these, smart products can offer innovative forms of customer value. This becomes comprehensible when considering the *product in use*: In combination with ubiquitous computing surroundings like in smart home applications, smart products adapt to preset preferences and user behavior. With adaptive *system integration,* these products access *product-related smart services*. In this way, the smart product is the tangible platform for a variety of services used depending on situation and context. Smart products can also be composed modular, giving the chance to extended functionalities if needed. *Modularity* allows the adjustment of products with regard of the users' preferences.

The inclusion of smart products into the product portfolio offers companies multiple benefits. First, the use of cyber-physical systems is not just for the advancement of the production process itself but also for the manufacturing of products with innovative forms of customer value. Second, with smart products it becomes easier to gather data about the product in use, which is valuable for the application fields of quality management and research and development.

4.5 Product-Related Smart Data

Just like in the sphere of industrial smart data, *product-related smart data* needs to be evaluated by an analytical process. As well as in the industrial process, the following application fields are preconditions for the derivation of valuable information: *Data warehousing, process engineering for data analysis, data analysis* and *data enrichment*. The outcomes of the data processing are used for two purposes. On the one hand, it is an enabling element for product-related smart services, on the other hand it enters the industrial value-adding process by being integrated as data from another context in the process of industrial data enrichment. Synonymous to industrial data processing, the product-related counterpart is dependent on reliable *cyber security* solutions.

4.6 Product-Related Smart Services

Product-related smart services constitute as the intangible part of the hybrid value creation complementing the tangible part, the smart product. In this context, c*onsumer service systems* act as a content aggregator, combining several independent services to a service package which suits to the individual needs determined by the consumer and the usage scenario. In most cases, these consumer service systems are controlled via apps installed from *app stores* on smartphones or other smart products. *User communities* can be used to gain information about user perceptions and usage behavior as well as to foster user driven innovations expanding the function ability of smart products and services. Another application field is the a*fter sales support* offered by the product manufacturer. With live support, customer service can provide assistance in case of functional problems. Software updates enable a continuous implementation of improvements coming from findings of smart data analysis of both industrial and product-related origin.

4.7 Utilization of the Application Map

In conclusion, a broad variety of application fields for industrial cyber-physical systems as well as their mutual influence on each other becomes apparent. Once more, it is to emphasize how cross-linked and interdependent the various application fields are. To give a complete overview of all application fields and related domains within this section, they are displayed in Fig. 2.

The reasons for the introduction of cyber-physical systems in the industrial value creation process are manifold: First, it offers the chance for further process efficiency with higher output and lower non-rectifiable rejects. Second, in many markets the customer demands have oriented towards individualizable products equipped with features pooled under the term smart as described in Sect. 4.4 of this chapter [23]. Often, for manufacturing these products the application of cyber-physical systems is a requirement. With an optimized production and an improved value proposition the own market position can be strengthened. Third, cyber-physical systems and the new level of data availability can give the basis for new business models and therefor an extension of companies' service spectrums or a repositioning on the market [27]. Summarizing, whether triggered by technology push or market pull and whether updating existing or building new structures, the introduction of cyber-physicals systems holds out the prospect of improvement of business success.

The decisive factors in this context are which application fields to choose, where to start, and how to proceed. Besides the aim to give a comprehensive overview of application fields for cyber-physical systems in the industrial context, the application map of this chapter is designed to support decision makers confronted with the stated above questions. How to use the map in a systematical way is described in the following.

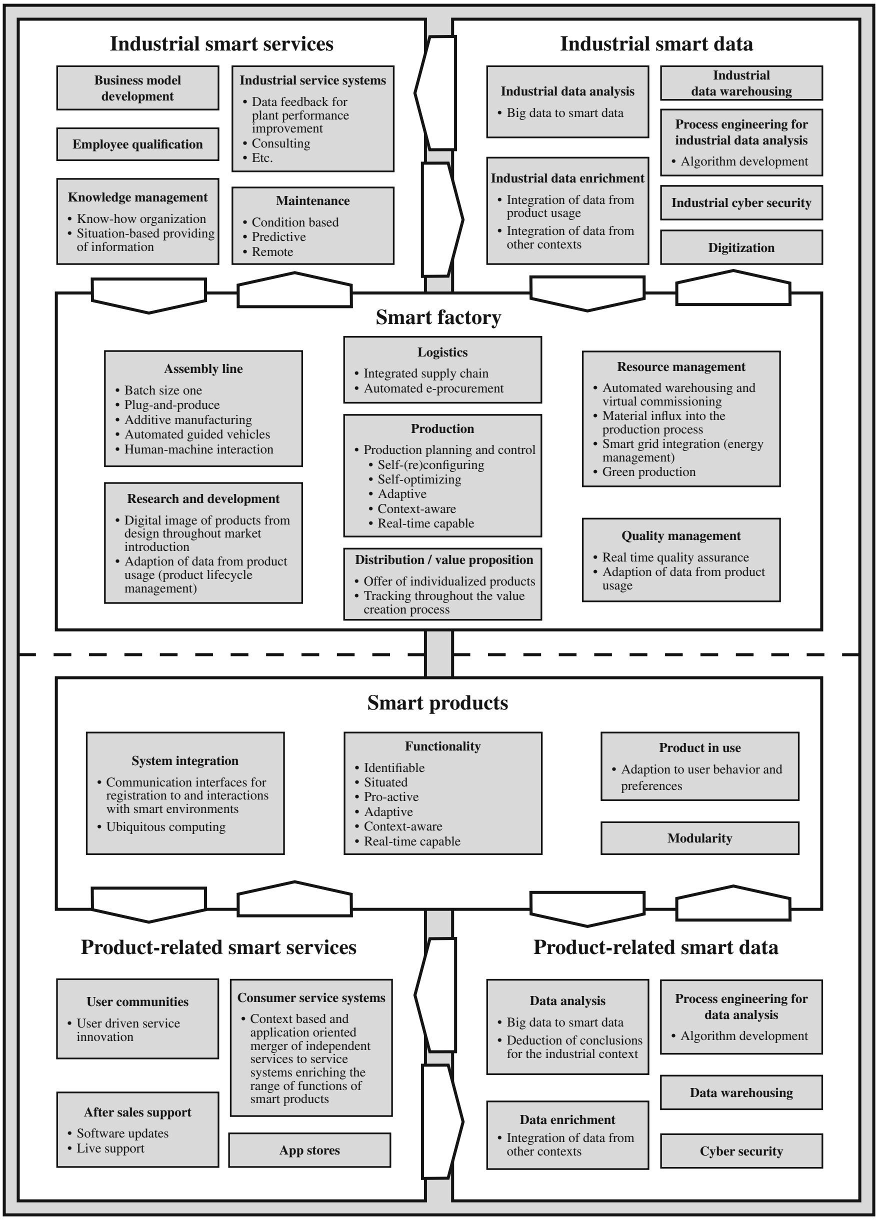


Fig. 2 An application map for industrial cyber-physical systems

Depending on the business scope of the company, a suiting sphere of the application map needs to be chosen. For companies with core competencies in the manufacturing process this is the sphere of the smart factory, for IT companies the spheres of smart data and for service providers the spheres of smart services. Once the suiting sphere is selected, specific applications fields within this sphere need to be chosen based on the individual companies' characteristics. These can be fields with pronounced expertise to strengthen but also fields of concern with potential for improvement. Then dependencies on surrounding application fields as well as potential synergy effects need to be estimate and anticipated. As the following step, again dependencies and synergy effects need to be estimated but this time not on field but sphere level. For example, an improvement within the field of maintenance is depending pronouncedly on the field of knowledge management within its own sphere (industrial smart services) but also on the fields assembly line, production and industrial data analysis from neighboring spheres (smart factory and industrial smart data). Depending on competencies, relevance for the business model and capital availability the decision needs to be made between in-house solutions or recourse on external service providers. This process should be repeated for every aimed application field with iterative cycles until the intended application scenarios for cyber-physical systems are planed satisfactory. During the process of implementation the map can be used for orientation and tuning continually. Once the implementation is done the map can serve as an underlying structure for validation and benchmarking.

The application map supports the decision making process on several levels, showing opportunities to improve and expand the own value creation concept with scopes for the establishment of value-adding networks with short term or strategic business partners. In this process the map is especially helpful due to the comprehensive view it gives on the implementation of cyber-physical systems in form of a holistic framework both on technological as well as on managerial level. Supporting this, the elaborated categories of Sect. 3 give a good orientation in which general topics expertise is needed for the professional handling of industrial cyber-physical systems.

5 Summary and Outlook

In this chapter, the foundations of cyber-physical systems were looked at in different dimensions. The organizational dimension was identified as most critical for the further development in the field. The categories in which improvements can be expected in the future were discussed and displayed in detail. There are nine categories with different scopes but all relevant and necessary for various applications of cyber-physical systems. Finally, concrete fields of application for the implementation of cyber-physical systems to reach such improvements were named and categorized and linked among each other. The application map is expected to help decision makers in the process of identifying suitable application fields for industrial cyber-physical systems and then implementing them into these matching to their business situation.

Due to the dynamic development of the field and the large research and development funding on offer, the future direction of cyber-physical systems is hard to foresee. The ability of managers to gain orientation about the possibilities for technical progress and the opportunities for business success will play a decisive role. The application map introduced in this chapter is only a starting point for providing research-based support for the extensive implementation and fruition of potentials of cyber-physical systems in industrial value creation processes.

References

1. acatech (ed) (2011) Cyber-physical systems-Innovationsmotor für Mobilität, Gesundheit Energie und Produktion., acatech POSITIONSpringer, Heidelberg
2. Ali ABMS, Azad S (2013) Demand forecasting in smart grid. In: Ali ABMS (ed) Smart grids—opportunities, developments, and trends. Springer, London, pp 135–150
3. Bock T, Linner T, Ikeda W (2012) Exoskeleton and humanoid robotic technology in construction and built environment. In: Zaier R (ed) The future of humanoid robots—research and applications. InTech, Rijeka, pp 111–146
4. Böhmann T, Leimeister JM, Möslein KM (2014) Service systems engineering—a field for future information systems research. Bus Inf Syst Eng (BISE) 6(2):73–79
5. Brecher C, Jeschke S, Schuh G et al (2011) Integrative Produktionstechnik für Hochlohnländer. In: Brecher C (ed) Integrative Produktionstechnik für Hochlohnländer. Springer, Berlin, pp 17–82
6. Bundesministerium für Bildung und Forschung (ed) (2014) Industrie 4.0-Innovationen für die Produktion von morgen. BMBF
7. Byoungsoo K, Minhyung K, Hyeon J (2014) Determinants of postadoption behaviors of mobile communications applications: a dual-model perspective. Int J Hum Comput Interact 30 (7):547–559
8. Cyphers.eu (2014). http://www.cyphers.eu/sites/default/files/D2.2.pdf. Accessed 14 Apr 2015
9. Fallenbeck N, Eckert C (2014) IT-Sicherheit und Cloud Computing. In: Bauernhansl T, ten Hompel M, Vogel-Heuser B (eds) Industrie 4.0 in Produktion, Automatisierung und Logistik. Springer, Wiesbaden, pp 397–431
10. Fortino G, Guerrieri A, Russo W, Savaglio C (2014) Middlewares for smart objects and smart environments: overview and comparison. In: Fortino G, Trunfio P (eds) Internet of things based on smart objects—technology, middleware and applications. Springer International Publishing, Heidelberg, pp 1–27
11. Geisberger E, Broy M (eds) (2015) Living in a networked world—integrated research agenda cyber-physical systems (agendaCPS). acatech Study. Herbert Utz Verlag, Munich
12. Gutierrez A, Dreslinski RG, Wenisch TF et al (2011) Full-system analysis and characterization of interactive smartphone applications. In: IEEE international symposium on workload characterization, Austin, 6–8 November 2011, pp 81–90
13. Heinrich B, Linke P, Glöckler P (2015) Grundlagen Automatisierung-Sensorik, Regelung, Steuerung. Springer, Wiesbaden
14. hvm.catapult.org.uk (2016). https://hvm.catapult.org.uk/. Accessed 17 Apr 2016
15. Jeschke S, Vossen R, Leisten I et al (2014) Industrie 4.0 als Treiber der demografischen Chancen. In: Jeschke S, Isenhardt I, Hees F, Henning K (eds) Automation, communication and cybernetics in science and engineering 2013/2014. Springer International Publishing, pp 75–85
16. Lee EA (2008) Cyber physical systems: design challenges. In: 11th IEEE international symposium on object/component/service-oriented real-time distributed computing, Orlando, 5–7 May 2008, pp 440–451
17. Levin SL, Schmidt S (2014) IPv4 to IPv6: challenges, solutions, and lessons. Telecommun Policy 38(11):1059–1068

18. Liu SX (2016) Innovation design: made in China 2025. Des Manage Rev 27(1):52–58
19. Mahnke W, Leitner SH, Damm M (2009) OPC unified architecture. Springer, Heidelberg
20. manufacturing.gov (2015). http://www.manufacturing.gov/advanced_manufacturing.html. Accessed 30 Sept 2015
21. Manzei C, Schleupner L, Heinze R (eds) (2016) Industrie 4.0 im internationalen Kontext-Kernkonzepte, Ergebnisse, Trends. VDE, Berlin
22. Marwalder P (2011) Embedded system design: embedded systems foundations of cyber-physical systems. Springer, Netherlands
23. Mühlhäuser M (2008) Smart products: an introduction. In: Mühlhäuser M, Ferscha A, Aitenbichler E (eds) Constructing ambient intelligence. Springer, Berlin, pp 158–164
24. Nist.gov (2013). http://www.nist.gov/el/upload/Exec-Roundtable-SumReport-Final-1-30-13.pdf. Accessed 14 Apr 2016
25. Oks SJ, Fritzsche A (2015) Importance of user role concepts for the implementation and operation of service systems based on cyber-physical architectures. In: Innteract conference, Chemnitz, 7–8 May 2015, pp 379–382
26. Pflaum A, Hupp J (2007) Auf dem Weg zum Internet der Dinge-das Versprechen innovativer Smart-Object-Technologien. In: Bullinger HJ, ten Hompel M (eds) Internet der Dinge. Springer, Berlin
27. Porter ME, Heppelmann JE (2014) How smart, connected products are transforming competition. Harv Bus Rev (HBR) 92(11):64–88
28. Purser S (2014) Standards for cyber security. Best Pract Comput Netw Def Incid Detect Response 35:97–106
29. Reichwald R, Piller F (2009) Interaktive Wertschöpfung-Open Innovation, Individualisierung und neue Formen der Arbeitsteilung. 2, vollständig überarbeitete und erweiterte Auflage. Gabler, Wiesbaden
30. Sommer L (2015) Industrial revolution—industry 4.0: are german manufacturing SMEs the first victims of this revolution? J Ind Eng Manage (JIEM) 8(5):1512–1532
31. Thiel C, Thiel C (2015) Industry 4.0—challenges in anti-counterfeiting. In: Reimer H, Pohlmann N, Schneider W (eds) ISSE 2015—highlights of the information security solutions Europe 2015 conference. Springer, Wiesbaden, pp 111–120
32. Tolio T (ed) (2009) Design of flexible production systems: methodologies and tools. Springer, Berlin
33. VDI (2013) Thesen und Handlungsfelder-Cyber-Physical Systems: Chancen und Nutzen aus Sicht der automation
34. Velamuri VK, Neyer A-K, Möslein KM (2011) Hybrid value creation: a systematic review of an evolving research area. J für Betriebswirtschaft (JfB) 61(1):3–35

Cyber-Physical Electronics Production

Christopher Kaestle, Hans Fleischmann, Michael Scholz, Stefan Haerter and Joerg Franke

1 Trends and Requirements in Modern Electronics Production

The industry of electronics production is driven by miniaturization, function integration, quality demands and cost reduction. This led to highly automated rigidly linked production lines dominated by surface mount technology (SMT). The following section illustrates the technological possibilities pushing new product and production developments as well as economic demands behind the need for more flexibility. Consequently, the necessity of a cyber-physical electronics production is derived and embedded into suitable logistics and production concepts.

C. Kaestle · H. Fleischmann · M. Scholz · S. Haerter · J. Franke (✉)
Institute for Factory Automation and Production Systems (FAPS), Friedrich-Alexander University Erlangen-Nürnberg, Egerlandstrasse 7-9, 91054 Erlangen, Germany
e-mail: joerg.fanke@faps.fau.de

C. Kaestle
e-mail: christopher.kaestle@faps.fau.de

H. Fleischmann
e-mail: hans.fleischmann@faps.fau.de

M. Scholz
e-mail: michael.scholz@faps.fau.de

S. Haerter
e-mail: stefan.haerter@faps.fau.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_3

1.1 Miniaturization and Function Integration

Cyber-physical manufacturing networks bear the chance to change the face of tomorrow's electronic and mechatronic products as well as their production systems. The classic production engineering is currently undergoing a major change due to the potentials of the transformation from automated production processes to smart production networks. Especially in the field of electronics production, automated and rigidly linked production lines are currently used. On the one hand, the quality and efficiency of the production lines up to factory or even enterprise level are monitored and evaluated in an integrated way. On the other hand, new requirements with increased complexity of the production place new demands for an optimized production with improved process control and innovative technologies.

Generally, the main process steps in SMT are the solder paste printing, the assembly of the components and the final reflow soldering process for electrical and mechanical interconnection. Initially, solder paste material is applied to the substrate materials by paste printing. The solder paste printing process is mainly characterized by a high degree of automation and a high throughput. In the following, the printed circuit boards (PCB) are populated with electronic components by at least one assembly machine. For efficient processing, the needed components are provided by the feeder sufficiently to the assembly machine in proactive quantities. In the final step, a mechanical and electrical interconnection of the electronic components and the PCB is achieved. The process control of this soldering process should ensure a sufficient temperature above liquidus temperature of the solder paste material at all interconnections and preferably low thermal stress to the components at the same time. Additionally, intensive inspection steps for process control are included, as illustrated in Fig. 1. Most commonly used is the solder paste inspection (SPI) for measuring the application of the solder paste and the automated optical inspection (AOI) after the reflow process. The integration of a further AOI step after the assembly process enables the holistic acquisition of quality data of the production. Automated x-ray Inspection systems (AXI) are used wherever defects need to be detected by non-destructive means.

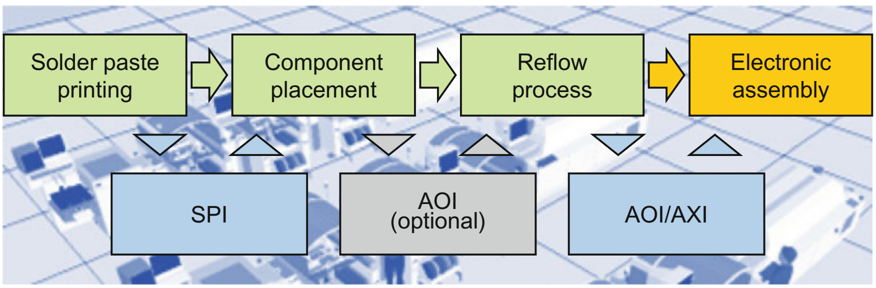


Fig. 1 Process chain in electronics production with optional inspection steps [16]

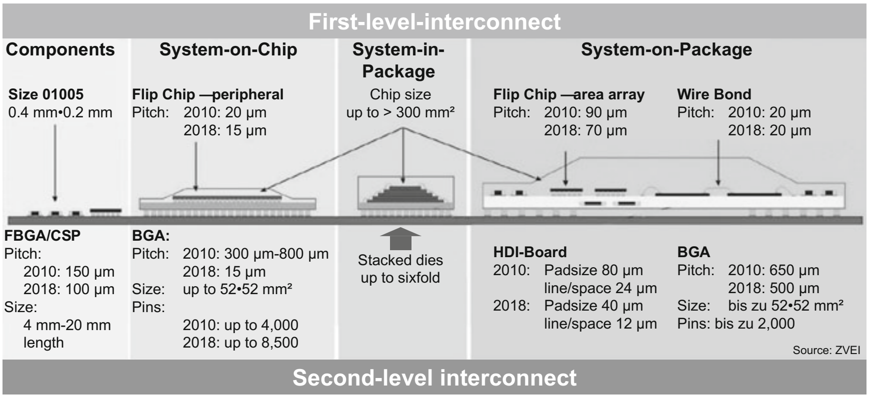


Fig. 2 Miniaturization as a key driver in the development of highly integrated systems in electronic devices [39]

With increasing miniaturization and complex features of modern electronic products, there is a growing demand for highly integrated printed circuit boards. The ongoing trend to miniaturized electronics has induced many developments towards size reduction and increasing performance in electronic products, as illustrated in Fig. 2. The market pull to this high integration initially focused developments in the component level, but can be found more and more in modern printed circuit boards (PCB). Small passive parts and highly integrated components for surface mounting provide smaller assemblies for mobile consumer products, medical applications, as well as sensor devices. Developments of system-on-chip (SoC), system-in-package (SiP) and system-on-package are the main drivers of First- and Second-Level-Interconnections of innovative packaging.

Through-silicon vias (TSV) have emerged to provide a highly integrated interconnection technique. Using TSV, applications with 3D integrated circuits and 3D packages can be produced. TSV provide high performance and functionality with highest densities. Another development is indicated by small packaging solutions and the embedding of passive and active components into printed circuit boards, e.g. the integration of RFID functionality in the inner layers of a PCB. This enables modern electronic products with improved electrical performance, high mechanical, thermal, and chemical protection, and high reliability. The introduction of these new components induces new requirements on multiple production processes and the used systems. Along with trends of ‘built to order’, ‘high mix, low volume’, and ‘one-piece-flow’ the complexity increases and leads to great challenges in electronic manufacturing. The targets of high efficiency with improved yield demand a deep process control for achieving high quality. As a first step, automation of production processes replaces manual processing. The use of close feedback control loop systems for consecutive production steps improves the achievable throughput and quality. The development for solutions of the technical diagnosis enable more complex knowledge-based expert systems.

The use of cyber-physical systems (CPS) enables the optimization of the production processes by innovative technologies and opportunities for advanced software systems. This replaces the established automation of the production processes and results in a self-optimized production network. CPS are characterized as self-describing systems with own intelligence, which dispose of an autonomous decentralized processing unit and are able to communicate directly over the internet [20]. The capability of self-learning and to adapting dynamically to the production environment enables the smart process control by manufacturing process integration.

1.2 Flexibility and Complexity

Today's electronics manufacturing is typically organized through an integrated production system (IPS) that controls manufacturing and logistics and includes interaction with suppliers and customers. Based on the three known industrial revolutions this resulted in the highly automated and highly efficient surface mount device (SMD) production concepts described in Sect. 1.1 that are most suitable for mass-production scenarios.

Frequently changing situations of demand, fluctuating input parameters and varying equipment availability represents a huge challenge for many electronics manufacturing companies as IPS are less suited for a quick and effective adaptation of production structures. In particular, the transformation of classical sellers' markets to modern buyers' markets requires sustainable measures for improving the flexibility to meet customers' demands. While the demand for customized products drives up the product and variant figures, decreasing product life cycles are recorded due to the increasing pace of innovation. This trend results in smaller lot sizes, more frequent product and version changes, and the demand for short throughput and setup times. Turbulent and dynamic changes in demand for goods as well as a lack of reliable sales forecasts require modern production systems for electronics manufacturing that allow a flexible response to different demand developments. The assembly of mass-customized products without an increase in product costs results is a great challenge with regard to handling the exploding complexity. Due to the increase in customer needs, flexibility and reactivity are more and more the factors of success. These changes are especially visible for small and medium sized electronics manufacturing services (EMS) that largely depend on day-to-day orders [7, 11].

Against the background of increasing flexibility demands, a significant rise in complexity accompanies these developments [14]. In this context, flexibility, with respect to inner and intra-production-site-mobility, gains importance. The increased complexity can often be observed in a drastic rise in inefficiency (muda, mura and muri) in form of waste, inconsistency, and overburden. In Sect. 2 various enablers and concepts are presented that allow for a production process with value-added action and minimal waste, thus facilitating a lean production through cybermanufacturing systems.

The complexity generated in a company by the aforementioned flexibility demands as well as the increased function integration in electronic products has become a critical cost factor and thus an essential issue for electronics manufacturing companies [35]. The causes for complexity are to be found within the company itself, and due to external factors. Whether a company can bring the exogenous complexity of the market and the resulting endogenous internal complexity in coexistence is not only a cost but also a strategic success factor. The increasing internal and external complexity is often justified by intimate customer and market involvement. With increasing competition, the search for a technological niche is often pursued. As customers are no longer willing to pay a price premium on volume products but demand mass-customized products, manufacturing companies try to offer an increased number of variants.

The increasing complexity has a considerable cost to that of a company's influence. The expected additional variants' higher contribution margins are often more than offset by increased complexity costs. Cyber-physical production systems bear the chance to break through this vicious cycle by facilitating flexibility at no extra complexity costs as well as the automation of overhead processes. They enable an electronics manufacturing company to run its production system at the sweet spot between flexibility, complexity, and cost efficiency.

Beyond the demand for flexibility in an electronics production system is the requirement for mutability [38]. This idea describes the ability of a production system to adapt its structures actively and quickly to changing and unpredictable tasks. These include in particular requirements for a product and variant flexibility, scalability, modularity and process flexibility in addition to compatibility and reusability of an electronics manufacturing system. The choice of the "right" degree of flexibility and adaptability is thereby a key challenge [28]. The right balance between additional expense and additional benefits from increased flexibility and mutability determines the economic efficiency of the production system. The problem of rising variant numbers, falling batch sizes in combination with decreasing product life cycles and fluctuating input parameters is illustrated in Fig. 3. This environment is difficult to control with classical integrated production systems and assembly lines. In this economic environment cyber-physical electronics production systems bear the change to offer the necessary flexibility while keeping the complexity in line.

Thus, the idea of Jidoka (自働化, Autonomation), automation with human intelligence is brought one step further to automation with human and machine intelligence whose interaction will be discussed at the end of this chapter. In doing so, the assembly of individual electronic products at the cost of mass-production can be established.

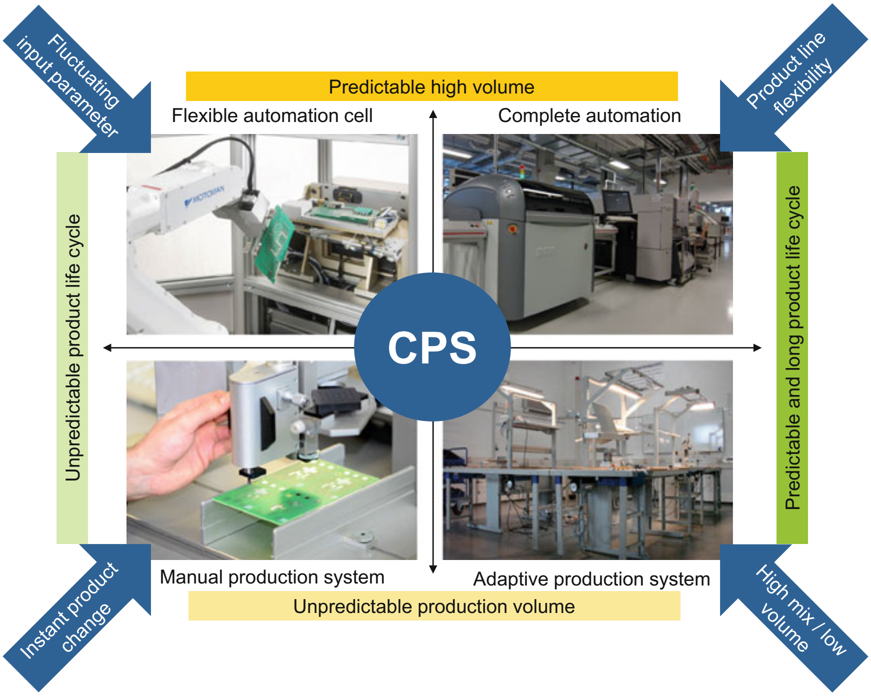


Fig. 3 Flexible production at mass-production efficiency enabled by CPS

1.3 Logistics and Production Concepts

The quota of intralogistics processes on the whole through-put time in electronics production is often underestimated. Like shown in Fig. 4 the transportation and waiting times over the whole production process add up to over 90 %. This indicates the high rationalization potential of common intralogistics solutions.

Current material handling in electronics production: A manufacturing system is the entire components that are necessary for converting a workpiece from one state to another [26]. Accordingly, an electronics production system consists of a great amount of subsystems. The specific tasks of these subsystems are generally related to the area of material flow systems or information systems. Material flow systems connect the physical parts of the manufacturing process such as machines, manual work stations, warehouses as well as transport and handling systems. Information systems include the immaterial objects of the material flow such as data or control algorithms, which are necessary for organizing and controlling the manufacturing process [26, 33, 36].

Depending on the spatial and organizational structure of the manufacturing site, common production systems are divided into three essential principles [6]: Line production, batch production and job shop production, as seen in Fig. 5.

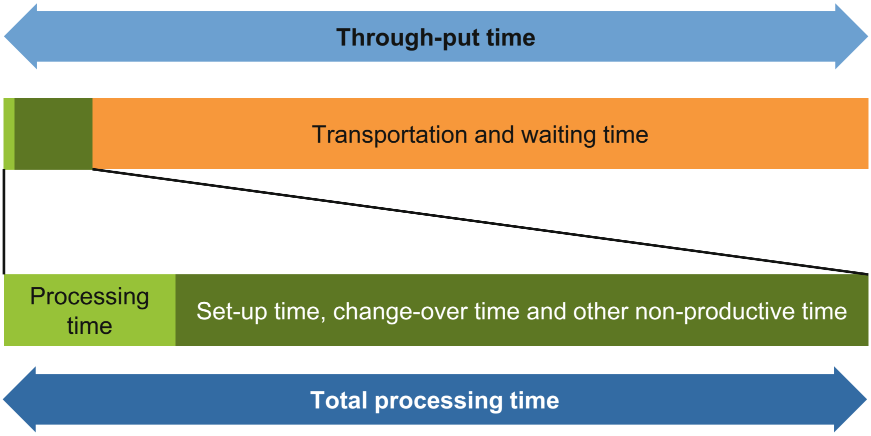


Fig. 4 Qualitative ratio of intralogistics processes on through-put time

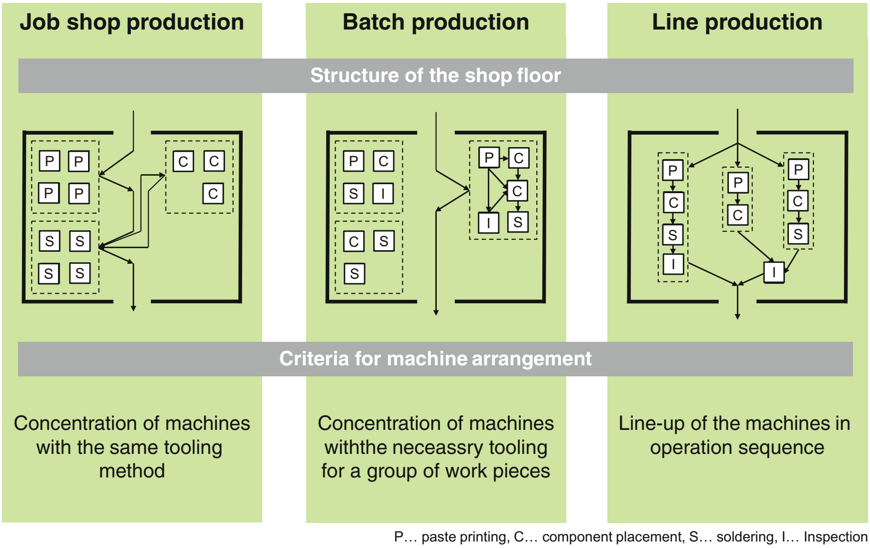


Fig. 5 Principles of machine arrangements

The basic principle of line production is used when an SMD electronics production facility has an in-line structure. In the current production environment this concept of SMD production is the most common organizational structure. In the in-line structure concept the particular manufacturing units are integrated and directly connected to each other with a continuous conveyer (see Fig. 6). This leads to a fixed connection of the manufacturing units, which results in a rigid process organization and hence short throughput times for the whole system [4].

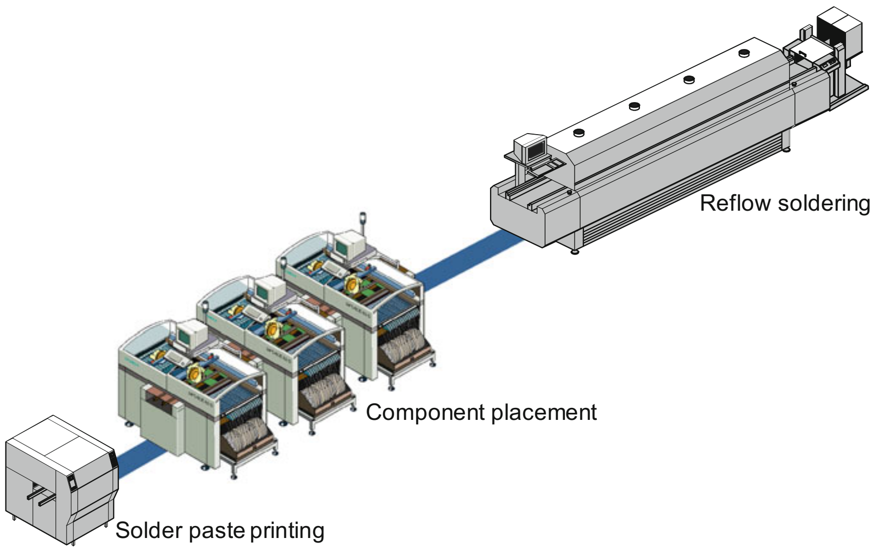


Fig. 6 In-line structure of an SMD assembly line

Furthermore, the line production principle leads to a clear and structured material flow, short work in progress and low transportation costs. This form of organization is especially suitable for workpieces of high quantity and low variance [6].

Within an on-line structure (Fig. 7 left) the particular manufacturing units are linked to each other with a central conveying system to increase the flexibility of an SMD production system. In this principle deflectors enable the system to transfer the workpieces between the production lines. The transported circuit boards can switch the lines between two operations. Due to this higher flexibility compared to the in-line structure, it is possible to produce a higher amount of variants. Furthermore, this approach reacts more flexible to a variation in the lot sizes within mass-production. However, this principle requires a complex and expensive

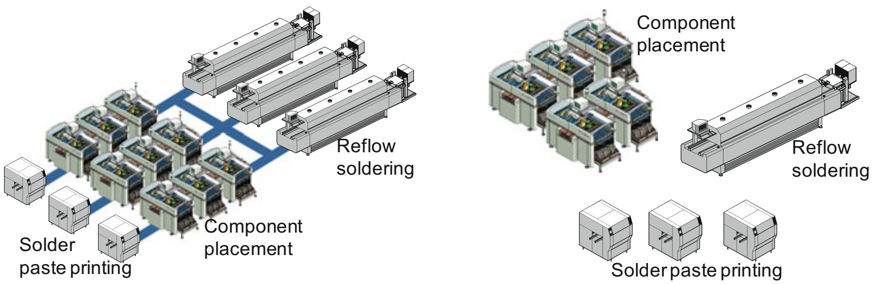


Fig. 7 Online structure (*left*) and offline structure (*right*) of assembly lines in electronics production

material flow system and an exact-levelled production control to achieve high utilization rates of the production units [31].

The third common principle of electronics production is the off-line structure (Fig. 7 right). Here, the manufacturing units are placed as stand-alone machines or combined as manufacturing cells with similar machine types according to the batch production layout. These units are not connected to each other with a fixed transport system. The transport of the circuit boards is mostly realized by hand. Therefore, the offline structure allows a more flexible, operation-oriented material flow within an electronics production site [21]. The domain of this principle is the manufacturing of small lot sizes with variance between the lots, but it is accompanied with a high occurrence of manual handling of the parts. Nowadays, the offline structure is used at job-oriented production sites and rarely for SMD production [31].

Technology Push. The flexibility of production sites with individualized routing and pathing of goods is currently prevalent in small-batch productions of large-sized products. For example, manufacturing the intermediate case of an aircraft engine at MTU Aero Engines in Munich is accomplished with an automated guided vehicle (AGV) system. The driving concept here is the linking the synchronized stations of the final assembly to each other and the preassembly. The parts are transported with AGVs from the preassembly to their particular station of the final assembly. There the parts are installed into the housing. With this concept approximately seven modules are finished each week [37].

This example shows the typical use of a system with an individualized transportation of the parts through the production site because of the high acquisition costs of AGVs. The driving costs behind these kind of vehicles are the sensors, actor and the on-board processing units. However, a price reduction within the last years has indeed become evident. The price for 3D-vision systems, which were used for research and special industrial applications, has decreased from several thousand euros down to a couple hundred euros. The root cause for this is the miniaturization and functional integration as shown in Sect. 1.1 and the emergence of 3D-vision systems into the consumer market. Applications for video game consoles such as the Kinect have particularly reduced the production costs of these vision systems due to their proliferation. The same trend is visible in the field of LIDAR systems where the costs for industrial AGV applications are ten times higher than systems with the intended use of research. Also, ultrasonic range sensors for consumer robotic products for observing the immediate vicinity are acquirable for less than ten euros. Not only have the costs for sensor systems decreased within the last years but also single board computers (SBCs) have benefitted from their introduction into the consumer market. Applications like the Raspberry Pi, Arduino or the Udoo Board are used for small embedded systems due to their miniaturization and functional integration.

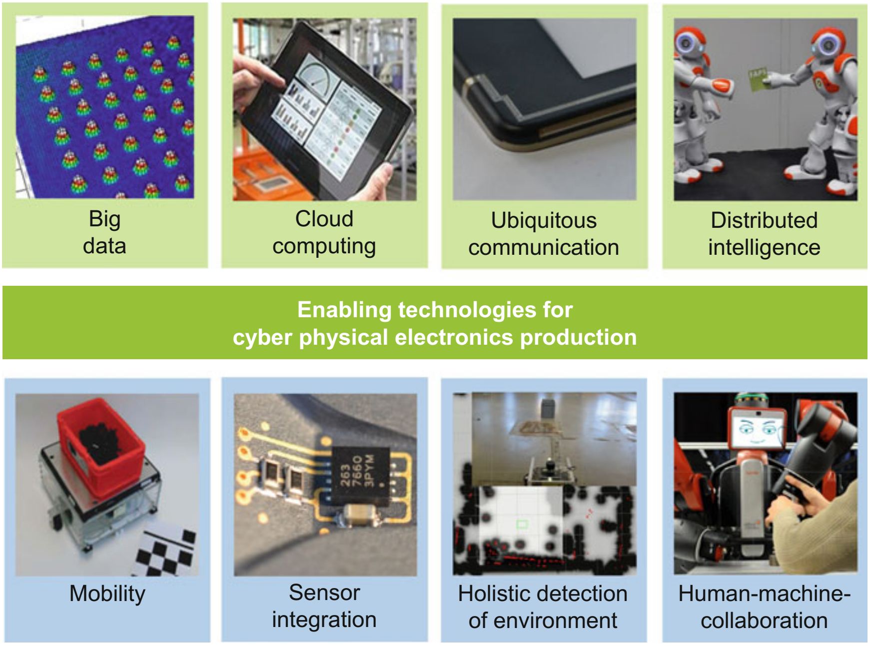


Fig. 8 Enabling technologies for cyber-physical electronics production

2 Enabling E-CPS Technologies

In order to face current trends and to meet the requirements of modern electronics production a variety of enabling technologies must be integrated into today's production environment. For this, an integrated production system can be promoted to a cyber-physical production system. The major technologies needed are shown in Fig. 8.

Technical solutions that facilitate ubiquitous communication, sensor integration, and a holistic detection of the environment will be illustrated in Sect. 2.1 from a hardware perspective. In Sect. 2.2 these enablers will be examined from a software point of view and complemented by requirements for big data, cloud computing, and distributed intelligence. Section 2.3 will focus on the technological integration of these enablers into a production environment, adding the need for mobility and human-machine collaboration.

2.1 Sensor Integration, Printing Technologies and Communications

Current trends in sensor and information technology enable the possibilities provided by Industry 4.0, to track and use all process data [34]. The main goal is the transition of this raised big data to become smart data as the effective and efficient use of that data is still not known. In 1993 one of the first ideas of closed loop applications are described in [9, 10]. The correlation between the solder paste printing and SPI for achieving better control about the printing results is highly offered in the industry. For example, a closed loop can be used for regarding a measured offset after the printing process by correction of the positions in the assembly of the components. In general, the use of advanced systems for SPI in an ongoing production generate a huge amount of data but a clear correlation of all inspection data is not obvious by itself. The main problems are the high demands for the data management and the growing gap between the needs for high performance of the inspection systems and increasing demands of the PCB. This is induced by the miniaturization of the components and higher integration using a smaller area on the PCB layout. The correlation of the collected data provides enormous potential for increasing the performance of a SMT production line. When all data of the manufacturing is tracked, the failure development for the whole process chain can be investigated. With increasing data base, the conclusions for process control are more statistical proven and can lead to a predictive manufacturing process by evaluation of the 'smart data'.

Miniaturization and new components mentioned in Sect. 1.1 enable the transformation of ordinary material, semi-finished products, transportation devices, and even machinery itself into cyber-physical systems. Additional to miniaturization of efficient electronic components, the embedding of components into printed circuit boards leads to smart packaging products. Besides the embedding of active and passive components, the integration of RFID can be exemplarily mentioned. Usually, a RFID device for automatic and contactless identification and localization requires an IC tag and an antenna. By using multi-layer circuit boards and high-frequency module techniques, antennas can be incorporated within the substrate. By this technology, PCBs are enhanced to be used in a smart production by accessing the information inside the product.

Furthermore, printing technologies such as ink-jet and screen-printing can be used for a flexible integration of printed sensor and communication elements on PCBs. An even more versatile technology for the integration of sensors and the enabling of ubiquitous communication is the aerosol jet printing (AJP) process. By printing versatile structures even of three-dimensional surfaces, this digital manufacturing technology can transform materials and semi-finished products into cyber-physical systems [19]. Figure 9 demonstrates possible use cases that can be achieved with AJP. Printing electronic components such as antenna structures shown in application example two may possibly be the most important feature. This creates a smart product by giving each material, component or semi-finished good the ability to communicate with its environment.

The AJP technology presented in Fig. 10 is a maskless and contactless, direct-writing technology, which can process a wide range of functional inks based on conducting as well as insulting materials [18]. The ink is pneumatically atomized inside the print head and the generated aerosol is carried to the virtual impactor. There, it is densified and subsequently guided to the printing nozzle. Inside the

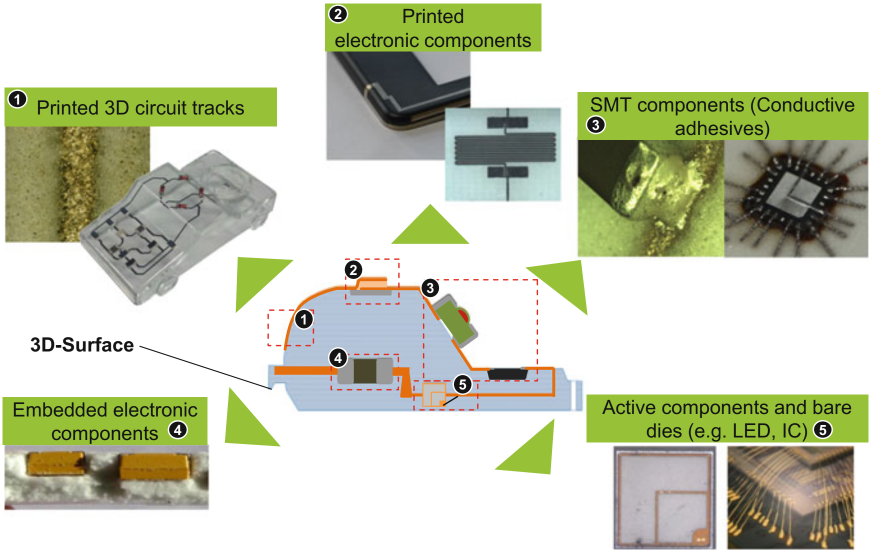
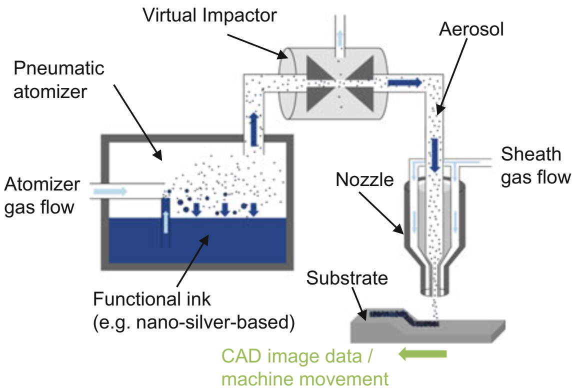


Fig. 9 Opportunities for the electronic functionalization of two- and three-dimensional objects by aerosol jet printing [19]

compact nozzle the aerosol is aerodynamically focused by an added sheath gas and is finally sprayed onto the substrate's surface. Depending on the processing parameters and the nozzle's shape, a line width of <100 µm can be printed. In addition, a focal length of the aerosol beam and a greater nozzle stand-off enable printing on complex 3D surfaces [17]. From an automation point of view, the aerosol jet process provides several advantages for printing functional structures on 3D substrates.

The custom-designed integration of sensor and communication technologies on machine, product, and component level is an essential enabler for the creation of holistic cybermanufacturing systems in electronics production.

Fig. 10 Schematic representation of the AJP process [15]



2.2 Software Systems

The goal of Industry 4.0 is to fundamentally improve industrial processes in the domain of production, supply chain management, and engineering. Products and equipment become intelligent objects or entities, so-called cyber-physical systems (CPS). Through the extensive interconnection of CPS, more efficient production processes can be achieved. CPS are able to communicate, to perceive their environment, to interpret information and to act on the physical world. These properties enable decentralized, autonomous smart factories with the ability for self-control and self-optimization. Functions of central information systems such as enterprise resource planning (ERP) or manufacturing execution systems (MES) are shifted to CPS, such that there is a gradual dissolution of the automation pyramid to an interconnected production grid [20] (Fig. 11).

In the context of smart factories, various approaches and solutions are currently under discussion in order to develop sophisticated production systems. These mainly include concepts and technologies such as the internet of things and services (IOTS), cloud solutions or agent systems that enable interconnection, communication, and data exchange of CPS in industrial domains. With the concept of the manufacturing service bus (MSB), principles of information technology and service orientation are applied, which serve as a framework for a successful implementation of versatile self-organizing structures [1].

The standardization of communication is playing a decisive role at this point in time. In current academic discourse, different communication standards such as open

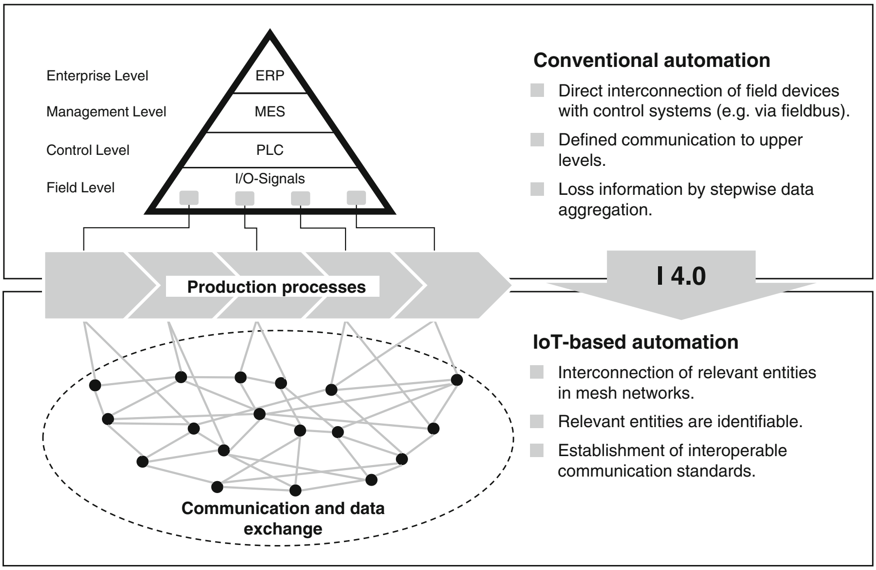


Fig. 11 Decomposition of classic automation pyramid towards ubiquitous communication [20]

platform communications unified architecture (OPC UA), MTConnect or message queue telemetry transport (MQTT) are being discussed. Each of these protocols has specific characteristics, capabilities, application domains, and backgrounds. For example, the standard OPC UA has its origins in industrial automation technology. The origin of MQTT is the IOTS and machine-to-machine (M2M) communication. This part focuses on essential requirements for communication and data exchange in Industry 4.0 as well as on the capabilities of established protocols [5].

Smart factories are currently in an early stage of research and development. In initial scientific publications, the terms 'Industrie 4.0' and 'cyber-physical production systems' (CPPS) appeared at the beginning of 2013. Although basic ideas and solutions had been published before, today there is a nearly implicit description of the requirements. The description is either at a high level of abstraction, declared as challenges or in the form of artifacts of design-oriented research. To extract requirements of Industry 4.0 from scientific publications, the method of qualitative content analysis according to Mayring [25] was used for the analysis. By defining clear rules and a systematic approach, this method allows reproducible and reliable results. The purpose of the analysis was the structured recording of demands on the communication and data exchange of industrial CPS in the context of smart factories. Based on scientific publications (journals, conference papers, and white papers) that address the topics of Industry 4.0, requirements were analyzed and derived by induction into a categorical system. The results of the analysis are 11 categories of requirements for communication and data exchange (Table 1).

Due to its scope, the issue of data protection and IT security is not considered in this part. With regard to the requirements for communication and data exchange, four basic cases of interaction in industrial CPS can be derived (Fig. 12): (1) transmission of data, (2) retrieval of data, (3) initiation of actions and (4) monitoring of the environment.

Within this system, information is transmitted on status, description, life cycle or knowledge. Status describes the condition of a specific entity. Description entails all the necessary information that is necessary for the description of an entity. This includes bills of materials or routings as well as features and capabilities of the CPS. Life cycle data entails information on the properties and states in development, production, and usage. Knowledge includes information from formal experience or expertise. The interaction takes place between different CPS (for example, intelligent products and equipment) and information systems in production. They include business application systems (such as ERP, MES), engineering systems (e.g. PLM, Digital Factory) or cloud-based applications. In addition, there are overriding requirements and assumptions of communication and data exchange. These include uniform and cross-system semantics, real-time processing, event control and the unique identification of entities throughout the IOTS.

Since the creation of the world wide web, a variety of standards, such as the hypertext transfer protocol (HTTP), have prevailed. The protocol family 'transmission control protocol/internet protocol (TCP/IP)' enables standardized communication between various entities on the internet. Likewise, in CPS, the standardization of communication and data exchange, as well as the definition of

Table 1 Requirements for communication and data exchange in industrial CPS

Category	Requirements
Engineering	– Synchronization of data in production systems with models of engineering
Cyber-physical systems	– Unique identification of entities in the IOTS – Connection and communication with other CPSs – Perceive, understand and interpret the environment – Awareness and monitoring of individual conditions – Autonomous triggering of actions
Flexibility and transformability	– Components and systems of different manufacturers, platforms and degrees of automation communicate through uniform standards – Components and systems know their own characteristics and capabilities and are able to communicate them – The adaption of entities occurs autonomously according to the environment
Interoperability	– Uniform semantics, technical guidelines, functions and states – Generic description, derivation and aggregation of information
Information models	– Information models represent physical production entities – Information models are updated as event-based – Representation of states, descriptions, life cycle data and knowledge
Real-time aspects	– Real-time requirements regarding communication, data supply, data processing and control
Comprehensive cross-linking	– Value-added comprehensive connection, communication and data exchange – Wired and wireless connections
Decentralized decisions	– Autonomous decision finding based on environmental conditions, superior goals, margin of discretion and predictable system conditions
Event-based decisions	– Self-reliant reaction due to unplanned events in production – Event-based interaction patterns, target definitions and freedom
Condition monitoring	– Monitoring and diagnosis of process data – Feedback and processing of process data for decision processes – Autonomous execution of actions e.g. maintenance
Knowledge processing	– Transfer of information and data in knowledge – Feedback of knowledge in decision processes – Autonomous usage of knowledge for self-optimization

necessary interaction mechanisms or communication models, plays a crucial role. CPS require open communication standards that allow the integration of new and existing information systems and entities [22].

While hard real-time requirements of determinism and response times take place at the field level (plant control systems, sensors, actuators), novel communication in industry 4.0 is classified more in the area of soft/near real-time requirements. Here, there are approaches to integration topologies that allow a logical or physical separation of industrial data transfer for classical automation systems. This communication is to be depicted on the IOTS, in which sophisticated, web-oriented

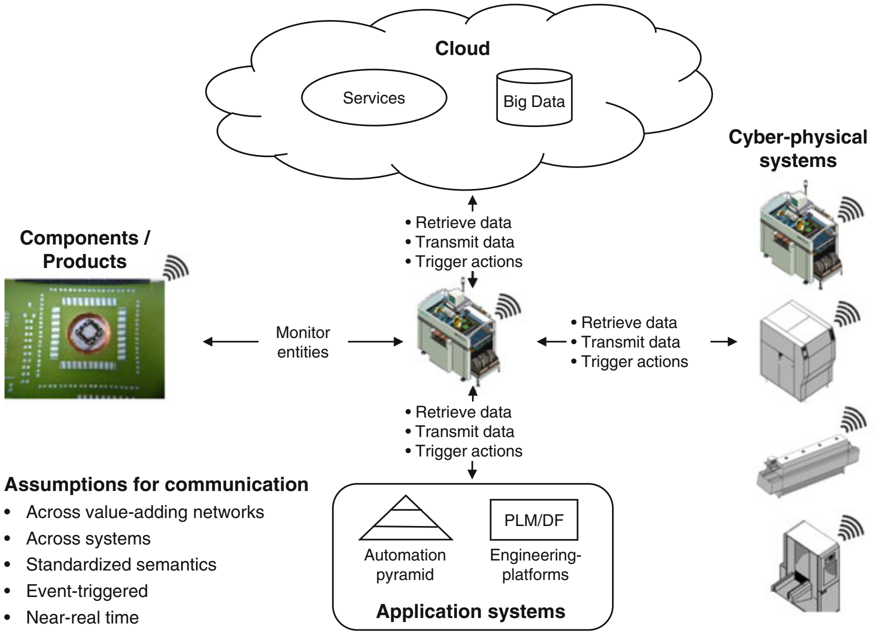


Fig. 12 Interaction between entities and systems in Industry 4.0

architectures and service or communication protocols are available. In addition to classic request-response procedures they provide new, effective mechanisms of interaction or communication models such as publish-subscribe or push-pull mechanisms. In addition, current research focuses on the development of reference architectures for CPS and the question of standardization [1].

For the comprehensive crosslinking, there are communication protocols available that have their origin in the world of general IT and possess potentials for smart factories. Representing this group here, MQTT is evaluated, which acts as a potential communication protocol in the IOTS. Since 2014, MQTT has been standardized according to the organization for the advancement of structured information standards (OASIS). An important feature is the focus on the publish-and-subscribe communication model. It is characterized by the existence of central broker for distributing information. Clients send messages that belong to specialized topics (publish). Other clients can subscribe to these topics on a message broker in order to receive messages (subscribe).

In addition to the protocols from the IT domain, there are other communication protocols, such as OPC UA or MTConnect, that are designed primarily according to industrial requirements. To elucidate this, OPC UA (object linking and embedding for process control unified architecture) will be discussed. Through the use of OPC UA, control hardware and field devices can be connected by means of standardized, interoperable and multi-vendor access methods. With OPC UA, a platform-independent, service-oriented architecture is available that has been well received and mentioned in a vast number of scientific articles. Unlike its predecessor OPC, OPC UA has a semantic, object-oriented data model. It supports minimal implementations for the integration of sensors and field devices into full-fledged applications such as cloud applications. Additionally, aspects of security and access rights can be established [5].

According to the OSI (open systems interconnection) layer model, a protocol is divided into seven layers (such as application, transport, or network layer). The transport layer is described through the established protocols (transmission control protocol) TCP and UDP (user datagram protocol), amongst others. With UDP, information systems can send so-called datagrams, which other processes or entities can receive. Classical and open Industrial Ethernet standards often use the connection-oriented TCP, which guarantees orderly and correct transmission and supports bi-directional links. The TCP/IP protocol suite, which enables communication on the Internet, is also one of the most important standards in industrial environments [22].

The requirements for communication and data exchange can be described with the four interactions (1) transmission of data, (2) retrieval of data, (3) initiation of actions and (4) monitoring of environmental conditions. These interactions can in principle be mapped with the push-pull, publish-subscribe or request-response communication models. The presented protocols also, in principle, support these communication models. Each communication model has specific strengths for the described interactions. Publish-subscribe is particularly relevant for monitoring entities and the environment. MQTT offers information about a mature, native publish-subscribe communication model, while OPC UA reflects the so-called subscription concept for this requirement. MTConnect sets web services to the machine level and utilizes the popular, request-response-based HTTP protocol. The requirement for standardized semantics allows interoperable communication and data exchange. The spectrum of transmitted information ranges from machine states via process parameters and routings to formalized knowledge. A standardized information model must be able to represent different information objects and levels and thereby be adaptable and expandable. Therefore, OPC UA provides a semantic information model that can be modified domain-specifically. In contrast, MQTT can integrate aspects of the semantics by the naming of the topics. Other protocols are based on the needs of specific industrial application domains, such as the communication standard, SEMI equipment communication standard (SECS), for the semiconductor industry [29]. The widespread demand for real-time capable data processing, networking, and control should be considered separately. In the field of automation technology, classical protocols and fieldbus systems allow real-time connections and defined response times. Contrarily, communication in Industry 4.0 is based on a global IOTS. To what degree real-time requirements, high transmission speeds and low response times are necessary is still unclear and the subject of current research. Furthermore, it is required that intelligent entities must meet production decisions autonomously. This requires the definition of rules and interaction mechanisms and the associated business logic. The implementation must be carried out on the CPS and be largely independent of the chosen communication protocol. Whether communication protocols have the potential to become standard for Industry 4.0 can only be determined upon fully specifying the requirements and interaction mechanisms between CPS and business IT systems.

Aside from the aspects of communication, big data applications are the enabler for cyber-physical production systems. According to a study by the McKinsey Global Institute, product development and manufacturing time can be reduced by up to 50 % by big data in manufacturing [23]. While a variety of process and product data is stored, the possibilities of data processing have yet to be sufficiently exploited. Some of the main enablers for smart factories are therefore new database types, as well as new approaches to software and hardware architectures for distributed computing. Relational database management systems (RDBMS) have dominated for years, alongside analytic databases (OLAP, online analytical processing). Currently, numerous, new NoSQL (Not only SQL) databases are emerging more and more. Based on the type of data, its inherent relationships, and the required scalability, users must choose between NoSQL database systems and conventional RDBMS. Key-value and column-family databases can be used when fast responses are required. Graph databases are superior when entity relationships are important. Document databases are able to cope with semi-structured data. Based on these databases, real-time analytic systems require instant access to the stored information to power advanced calculations on the status of machines, processes and parts, as well as historical data. This enables spotting unknown correlations between quality factors and influencing process parameters. Here, which set of parameters may lead to defective parts and how to avoid such states can be identified [1].

2.3 *Autonomous and Smart One-Piece-Flow*

In the future, intralogistics material flow systems must generate individual routings for each order. The sequence of the individual production steps and machines as well as the path through the shop floor must be generated individually for each workpiece. One decisive enabler for this scenario is that each circuit board can be carried and pathed individually similar to the concept of AGV in small-batch production.

On-board hardware of an autonomous and smart workpiece carrier (ASWC). The dimensions of such an ASWC in electronics production are deduced from the size of the handled load, the circuit boards. Therefore, its proportion is similar to a workpiece carrier of a common belt conveyer. There are various concepts for the driving and coupled axis to obtain the needed degrees of freedom. A prototype of an autonomous and smart workpiece carrier, which is shown in Fig. 13 (left), has a differential powertrain. It allows the system to rotate in place. To enable all needed degrees of freedom, the CAD model of the prototype (Fig. 13 right) shows a rotating transportation platform. The autonomous power supply of

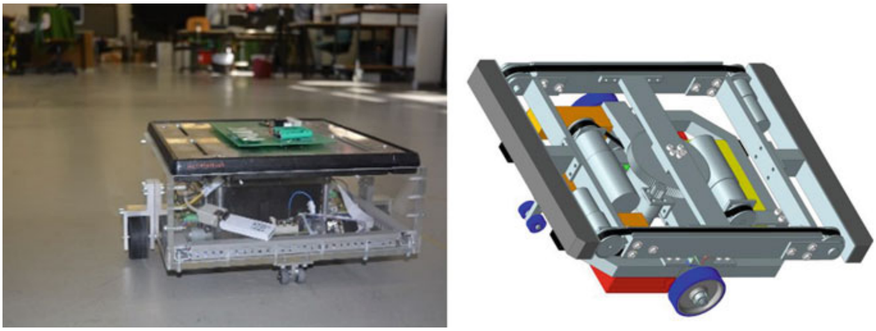


Fig. 13 ASWC with circuit board (*left*), CAD model of an ASWC (*right*)

such a system depends on the use case. Supercaps can be used for small operation areas and short distances between the charging stations. The advantage of supercaps is that they can quickly be recharged. Another possibility is the use of accumulators for autonomous power supply. Then only a small number of central caching devices is needed while the quantity of workpiece carrier increases. Additionally, idle time is used to recharge the whole fleet.

To ensure a safe system an ASWC needs sensors to digitize its surroundings. Therefore, various technologies are possible such as 2D and 3D vision systems, laser scanners or ultrasonic range detectors. What such a sensor concept can look like is shown in Fig. 14. Other indoor localization methods such as indoor GPS or the trilateration of Wi-Fi signals are either too expensive or lack the necessary accuracy to localize and route an ASWC. A common approach to digitizing the environment with a 2D laser scanner is the SLAM (simultaneous localization and mapping) method. A LiDAR (light detection and ranging) sensor measures the angle and distance to the obstacles in the system's surroundings and include this information in a map of the environment. This data can be merged with the data of the vision system. The structured light method projects a known pattern of light often in the infrared spectrum onto the scene. This pattern is deformed when it

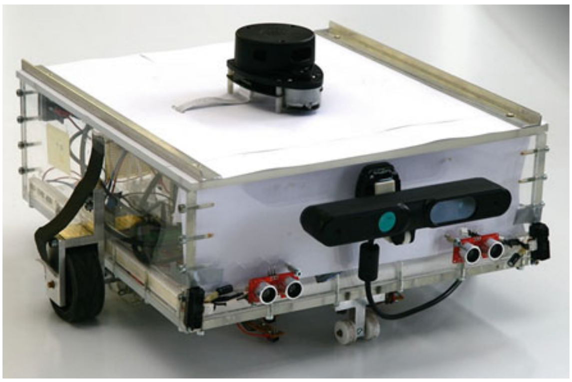


Fig. 14 ASWC prototype with on-board vision, laser and ultrasonic sensors and an embedded singe board computer (SBC)

strikes surfaces and objects. This allows the vision system to calculate the depth and distance of the objects. Another 3D vision concept is the ToF (time of flight) method which measures the time a light impulse takes to reach an object and back. These 3D vision systems are similar to laser scanners however with the advantage of being able to measure a whole scene at once. The main disadvantage is the range of vision compared to the LiDAR systems with 360° capabilities and the required computing capacity to stitch several 3D scenes to a panoramic view. Finally, low-cost ultrasonic sensors in the direction of motion can be used for a near-field obstacle detection and collision detection. If all other sensors fail, these sensors trigger an emergency stop directly to the motors of the ASWC [32].

A necessary hardware device for an autonomous and smart workpiece carrier is an integrated on-board computer. Thereby, a central computing device is not necessary. Therefore, the system is theoretical infinite scalability because each entity contributes its own computational power. ASWCs are thus ideal examples for cyber-physical systems. Each workpiece carrier is able to plan its path on its own, which allows an individualized transportation of the goods.

Concept for infrastructural sensors and locating of the carrier. Another approach to digitize the workspace of an ASWC is to generate a digital map with infrastructural sensors, for example ceiling cameras. This concept reduces the amount of necessary on-board sensors on the carrier since only near-field collision detection sensors are required. This concept is highly efficient to organize a big carrier fleet, which operates in a small workspace due to the breakeven point of the sum of on-board sensors and necessary infrastructural sensors. The accuracy of ceiling cameras depends on the ceiling height, the flare angle, and the resolution of the camera sensor. In typical industrial scenarios with a ceiling height of approximately five meters, a commercial, high-definition webcam possesses a sufficient resolution to navigate an ASWC. The infrastructural sensor system must detect different types of objects. Carriers, moving obstacles, standing obstacles and targets must be distinguished. Therefore, a computational device is needed to analyse the digital picture of the workspace. The use of one central device to stitch the images to one world frame is only possible if the amount of cameras is small. To follow the approach of CPS with distributed and embedded intelligence, a combination of the infrastructural sensors with a single board computer to one embedded system is more constructive (see Fig. 15). The SBC pre-processes the pictures directly on the camera device and only distributes the information that is needed for navigation as requested from an entity in the workspace. For example, the ASWC only needs the corners of an encircling rectangle to avoid collision with an obstacle. This concept of an embedded infrastructural sensor also reduces the amount of data to be sent via Wi-Fi. Furthermore, the distributed data from the sensor are usable for further intralogistics tasks such as navigation of the maintenance technician, digitalization and visualisation of the material flow in real time, the arrangement of machines and equipment [12, 32].

Navigation and path planning. After digitizing the workspace and locating the carrier and the targets, such as the pick-up and drop-off place of the goods, individualized navigation and path planning are necessary. The embedded intelligence

Fig. 15 Concept of embedded infrastructural vision sensor

of the carrier is able to calculate the path of the ASWC throughout the shop floor. As discussed above, the scalability of the whole system is provided because each system possesses path-planning capabilities, which obviates a single point of failure [12].

To path a circuit board through the production site, no perfect solution is needed. Probabilistic path planners are commonly used due to the computational power and the necessary runtime of the algorithm. Two different approaches are often used to solve such pathing, the generation of single-query trees (Fig. 16 left) and the generation of multi-query maps (Fig. 16 right). Pathing with a single-query tree is a combination of developing a solution tree and then searching for the best solution to connect the starting point with the target. The algorithm probabilistically defines points from the origin and connects them within a predefined range of the target. When the target is hit, the shortest path along the connections is calculated to traverse from the starting point to the target. The principle behind single-query trees functions independently of the tree generation from the starting point to target, vice versa, or a combination of both.

The other concept is the generation of a multi-query map within the whole workspace. Thereby, the connection points have a defined distance to each other. For each task, the specific starting point and target is inserted to the map and is connected to the nearest point. The task-specific solution is calculated along the connections. The generation of a single-query tree is always faster than the generation of a synonymous, multi-query map. With regard to the on-board and infrastructural sensor concepts, single-query trees are the best solution when all sensors are on board. Generating a multi-query map is advantageous when the workspace is digitized by infrastructural sensors. The position of all obstacles are known in real time and the map can be provided to the carrier. This leads to lower computational power requirements by the carrier. Routing each individual task becomes faster, because the map and the connections already exist and the task's individual starting point and target must be inserted before the path calculations.

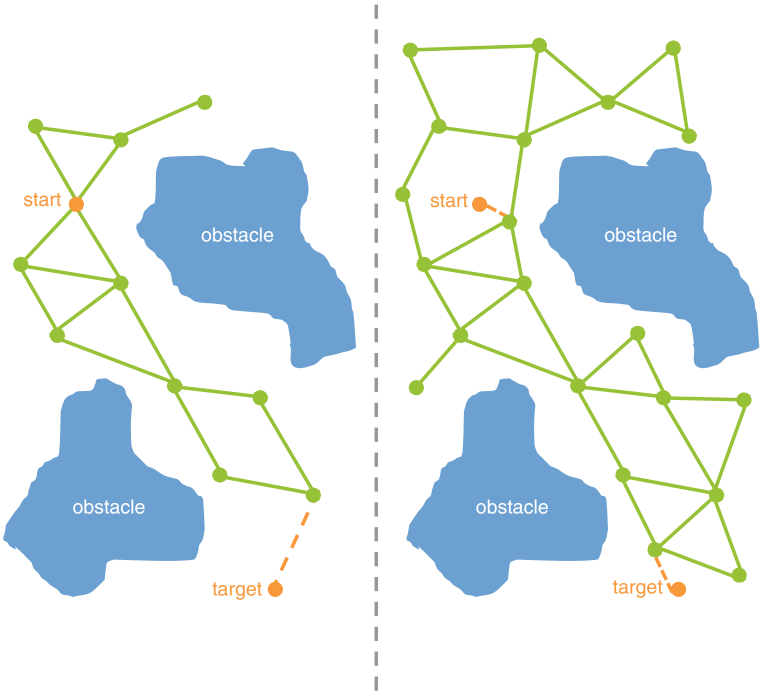


Fig. 16 Generation of a single-query tree (*left*) and a multi-query map (*right*)

3 Concept of a Cyber-Physical Electronics Production System

The enabling technologies in Sect. 2 of this chapter represent the basis for various shapes of cyber-physical electronics production systems. Big data technologies and cloud computing facilitate a self-learning electronics production line. Modern human-machine collaboration and ubiquitous communication between man and machine as well as machine and machine create a socio-cyber-physical electronics production. In a holistic approach the autonomous and smart workpiece carrier form the mobility basis for the physical, energetic, and digital connection of all entities. A sensor based all-encompassing detection of the environment controlled by global and local intelligence facilitates an integrated cyber-physical electronics production system. In this section the concurrence of these enabling technologies is illustrated in three different scenarios of cyber-physical electronics production systems.

3.1 Self-Learning Electronics Production Processes

The aim of a self-learning production line is the holistic integration of sensor-based data within the value chain of electronics production in an automated analysis and decision system for the recording and interpretation of accruing process and test data. Thus, improved production quality as well as increased process flexibility can be achieved across all process steps.

In its entirety, the electronics SMT production line is a strongly sensor-based system, which is subject to a variety of processes, systems, interfaces, and suppliers of large complexity. Due to the increasing complexity of production processes described in Sects. 1.1 and 1.2 process control is ascribed enormous relevance. This tremendous complexity increase with the progressive miniaturization of passive components and by processing highly integrated components continually presents new challenges. Since even the default of seemingly simple components can lead to the failure of full assemblies, quality levels of simple transistors is calculated in ppb (parts per billion), such that there is great potential for optimization here [3]. This potential, however, is far from being fully realized since the holistic storage and evaluation of extensively collected process and quality data have yet to occur. Both the volume of data as well as its diversity in terms of inhomogeneous file formats and data sources have thus far prohibited a holistic evaluation along the production line. This has additionally prevented a timely in-line analysis and processing of the data, whereby the technical and economic potential has remained untapped.

This continuous increase in data diversity and complexity of data is in particular due to the 80–90 % reduced cost of MEMS sensors in the last five years as well as the significantly increased amount of connected machinery and equipment [24]. However, the combination of modern sensor technology and automated data analysis is only at the beginning of its development, despite these impressive figures. This is likely to change due to technological progress described in Sect. 2.2.

Furthermore, the inspection and monitoring of critical processes often continues to take place manually. A prerequisite for the development and implementation of automated and connected techniques is the integration of machinery into control systems and cloud-based data systems. This requires standardized interfaces on the device side or flexible integration at the control level. This forms the basis for implementing automated fault detection and classification processes as well as the automated tracking of process parameters. The efficient use of collected process and quality data through smart data methods provides the basis for establishing a cross-process quality control loop as shown in Fig. 17. Instead of individual processes, the entire value chain will be considered and statistical methods are used for a holistic process optimization.

Thus, the achieved increase in yield and a corresponding reduction in errors and rework costs facilitate the, thus far, untapped economic potential. Manufacturing flexibility and efficiency are increased through quick, self-regulated adjustments to all production processes. By correlating various influencing parameters, an overall

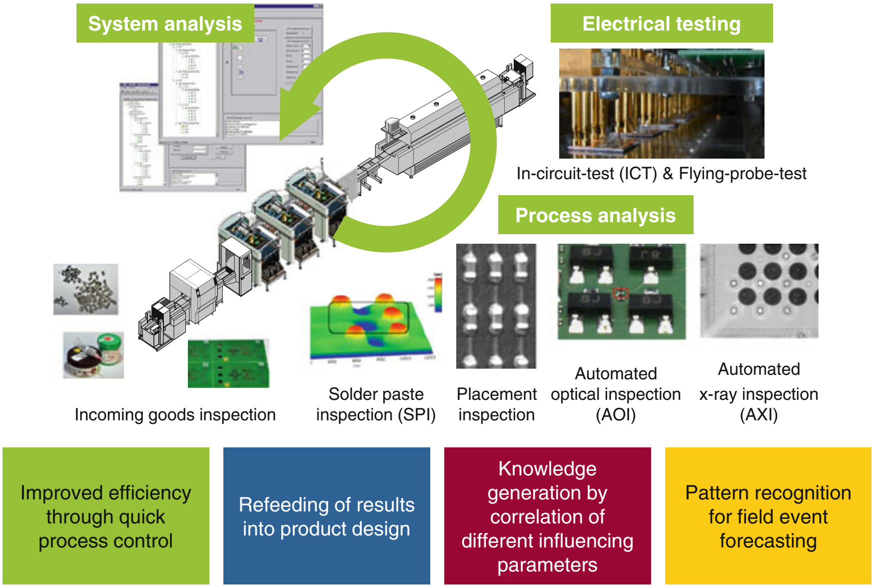


Fig. 17 Self-learning electronics production processes facilitated by big data technologies

understanding of the process can deepen, whereby insights can be passed on to the product design department. With the intensive use of knowledge intrinsic to date however inaccessible, both product quality as well as customer satisfaction can increase. The advanced knowledge of cause-effect relationships based on a holistic approach leads to an increase in first-pass yields as well as to increased product reliability. In addition, detailed forecasts of the failure behavior of critical components can be created through the collection and analysis of this data over longer periods.

In this context, safe, high-performance systems are needed to address these two temporally divergent problems—real-time, self-healing, in-line quality control, as well as the long-term recognition of cause-effect relationships. The database-systems must address issues of data organization and data management to create the bases for analyzing and forecasting changes in input variants and process parameters. Subsequently, a user-friendly presentation of data and correlations enables applications for decision support and the automation of control loops. In this context, the database system must be able to deal with challenges regarding the heterogeneity, accessibility, and usability of the data as well as issues regarding the quality and security of data. All these conditions and quality requirements for dealing with the resulting process, testing, and field data must be ensured on three levels. Firstly, in process, as has already been attained in a closed-loop process between a stencil printer and a solder paste inspection system. However, this must also occur across processes along the entire production line and more recently even across enterprises both at different manufacturing sites and throughout the entire product life cycle.

By making use of big data technologies and cloud computing, manufacturers are offered enormous potential to generate values from the diverse use of large sensor generated data volumes. The integration of data across the enterprise and the application of advanced analytical techniques help to increase productivity by improving efficiency, increasing flexibility, and promoting product quality [1]. Sophisticated analytics can significantly improve automatic decision making in production, minimize risks, and provide valuable insights, which would otherwise remain hidden. Sensor-based data provide the required raw material either to develop new algorithms or to use established smart data algorithms. Thus, new economic, environmental, and social opportunities arise. In developed markets, manufacturers can use the large amount of sensor-generated data to reduce costs and achieve greater innovation in products and services.

3.2 Assistance Systems

Novel assistance systems will ensure the operation of and human-machine integration in cyber-physical electronics production factories. The field of information and communication technology (ICT) and especially the web environment has been characterized by a high rate of innovation for the last ten years. Today's ICT systems are further characterized by high calculation speeds, vast memory capacities, and wireless network technologies. Additional features are miniaturized design as well as low costs, whereby their production is economically viable. This has led to highly integrated, powerful, and portable consumer hardware such as tablets and smartphones with worldwide distribution. These devices are capable of performing ambitious tasks concerning information acquisition and transmission. At the same time, the degree of maturity of web technologies (cp. HTML5, WebGL, SVG etc.) has improved so greatly that user-friendly and powerful web-based software tools can be developed. Browser-technologies such as HTML5 and JavaScript (JS) are used for platform-independent, configurable user interfaces on mobile devices [2, 27].

Web-based and mobile approaches distinguish themselves by numerous advantages in contrary to desktop-based systems for information representation in the field of production. There is ubiquitous, direct access to current information from diverse terminal devices without the need for installing application-specific software. In addition, the rollout and maintenance of software instances on client hardware can be omitted because the current version of an application is provided by a server during each usage of a system. Furthermore, transparent information representation is feasible over different platforms (Windows, iOS, Android) through a standard web browser. This is accomplished through a consistent web-based approach with the application of open-source web standards. No special adaption to hardware or operation system is needed [27].

With JS Engines, HTML5 technologies and the bidirectional web socket (WS) protocol, the web browser has established itself as a fully-fledged application platform (see Fig. 18). The layout of human-machine interfaces has been created as a combination of cascading style sheets (CSS) attributes and the hypertext markup language (HTML) structure. A cut between content (HTML) and design (CSS) has also been gained. HTML5 in its current version is a giant leap in evolution. HTML5 offers the integration of multimedia content without plugins such as Adobe Flash Player in addition to playback even on mobile devices. With the help of CSS, a website's layout is determined in the form of colors, font types, distances, etc. This contributes to a uniform presentation of websites within large projects. There are advantages to separating a layout into external files. The adaption of a website layout in a web browser is much faster in terms of application- or user-specific guidelines [30].

The objective of a worker information system (WIS) is an ergonomically correct and intuitively useful provision of the right information at the right place at the right time for manual assembly tasks. Thus, a WIS should assist a worker during the assembly of products with many variants and counteract the rising complexity of the worker's tasks. Customized products and the reduction in a product's life cycle due to rapid technological progress are reasons for this variance. The necessary flexibility in this field of production cannot be attained by automated assembly processes in principle and calls for the integration of manual and flexible assembly stations. The benefit lies in the immense cognitive skills of human beings in order to react to unexpected events, to independently plan further steps, to learn, to gain experience, and to communicate with other subjects. In order to maintain a competitive edge in a country with high incomes such as Germany, production-near employees must be qualified for and empowered to do their tasks through the application of information technology [13, 27].

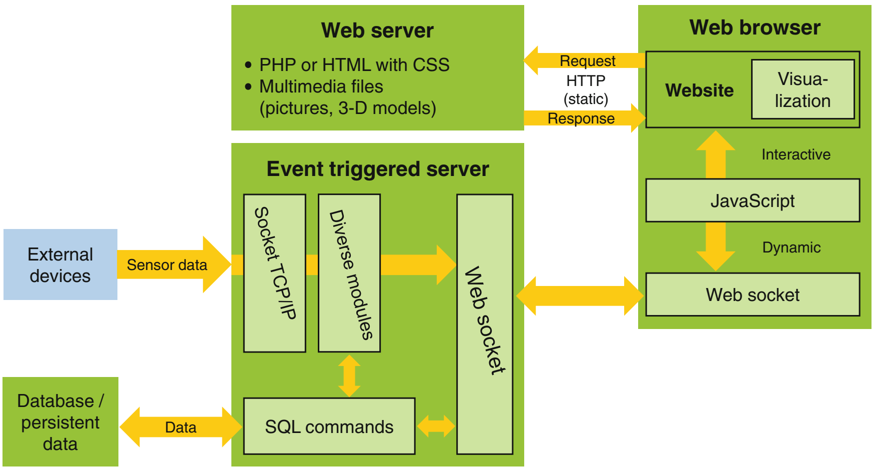


Fig. 18 Typical web application architecture

The main focus of a WIS must be directed at the worker as the central protagonist and critical factor for production success. The WIS must be effective and offer notably additional value for a worker in order to gain acceptance. At the same time, bidirectional data transfer from a WIS to a manufacturing execution system (MES) or a person responsible for production is mandatory in order to collect and exploit a worker's implicit knowledge. Thus, details regarding problems, solutions and improvements to efficiency can be provided to employees in product development, production planning and on the shop floor [13].

For a WIS, there is also a demand for media continuity as well as complete and reliable information provision in the form of order data, working instructions, bill of materials (BOM), quality testing information, drawings, or similar objects for a worker on the shop floor. The classical paper-based approach largely addresses previously existing data from enterprise resource planning (ERP) and production planning; hardcopies lack cycle-dependent and working-near orientation. This is also true for timeliness and for frequent changes e.g. at the start of serial production (media discontinuity) as well as the flexibility in representation. In addition, a worker has no overview of papers from different sources, which leads to his cognitive overload. Thus, a logical step in system evolution guides towards digital and intranet-based employee information systems with a usage or extension of established IT infrastructures, to which there are many advantages. Firstly, creation and maintenance processes are easier and faster. Secondly, this yields numerous opportunities to integrate multimedia files such as images, videos, audio and animated 3-D models into a WIS. As a consequence, WIS software inevitably uses web technologies [1, 13]. Figure 19 illustrates the combination of a virtual mockup and additional worker information.

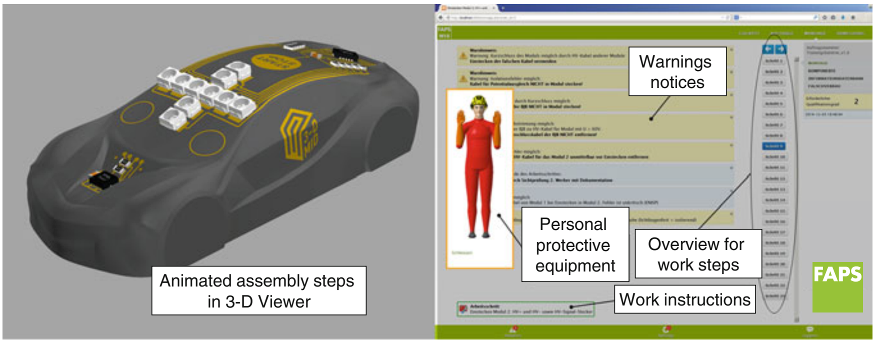


Fig. 19 Showing 3D product assembly models in TeamCenter Visualization Mockup (*left*) and further worker information on a website (*right*)

3.3 Integrated Cyber-Physical Electronics Production

The demands for the highest productivity rates with more than 100,000 placed electronic components per hour with an aspired error rate of only a few defect per million have led to the rigidly connected production lines described in Sect. 1.1 [8]. The transport of products and materials is accomplished via conveyor belts or large and inflexible driverless transport systems (AGV) that are described in Sect. 1.3. Against the backdrop of increasing product functionality and complexity, the importance of an increased flexibility of electronic production systems prevails. Currently, rigidly linked production lines prevent quick changes to the production sequence. This inadequacy is in particular visible during standstills of individual machines e.g. due to maintenance intervals, which results in a standstill of the entire line, and thus significantly impairs the overall equipment effectiveness (OEE). Another use case showing the inflexibility of current assembly lines is that of the reworking process. Rejected products are not discharged automatically from the manufacturing line and transported to a reworking station, since neither the transport systems nor the IT is able to do so.

This creates the possibility of a dynamic breakup of rigid manufacturing lines in electronics production using smart cyber-physical attributes. Thus, a dynamic, viably real-time, and self-organizing internal value chain can be developed according to different targets such as cost, availability, energy and resource consumption, flexibility and throughput time. To reach this goal, all enabling technologies presented in Sect. 2 must work hand in hand. This creates the possibility of a production setup as shown in Fig. 20.

The wider use of RFID technology or AJP antenna structures for mobile objects, such as electronic assemblies, device delivery systems in the form of rolls, trays and bulkcases, tools, stencils, solder paste and functional materials, enables ubiquitous communication and networking of all objects in the factory. Digital product memories are developed by which process parameters and information is stored directly on the workpiece. The use of communication standards such as OPC Unified Architecture guarantees a modern, efficient and interoperable connection of automation components, control hardware and field devices through a standardized multi-vendor access method. Currently, the production of RFID tags is realized by reel-to-reel (R2R) printing methods. Then, the RFID tags need to be applied onto the materials, machines, and products. Hybrid RFIDs applied by integrating discrete and printed features can create embedding solution in electronic components. Thus, autonomous communication between products, machines and transport systems in terms of material supply, maintenance, repair, machine set-up can be ensured at any time. Moreover, the additional use of cloud-based big data functionality enables the use of the collected data in the sense of “Operations Research” for the control of production, the production of statistical quality schemes and set-up optimization. The self-diagnostic capability of equipment can be translated as described in Sect. 3.1

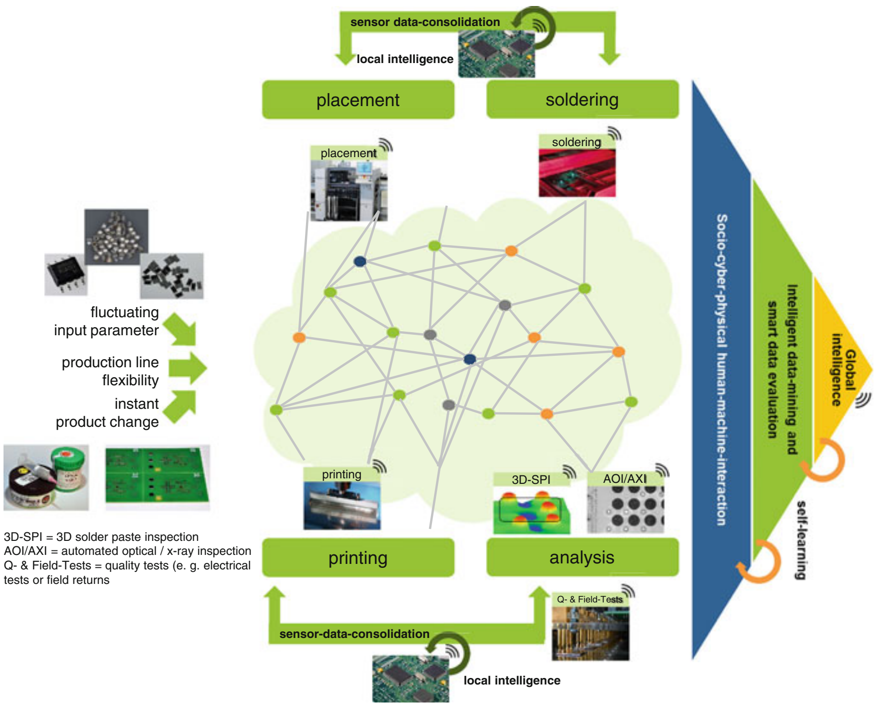


Fig. 20 Realization of a cyber-physical electronics production network by flexible connection of all entities

The basis of dynamically breaking up the rigid electronics production line concatenation is constituted by the "autonomous routing" of the semi-finished electronic modules in a truly one-piece flow concept through direct interactions between the workpiece, transportation systems and machines according to predetermined production strategies. Here, small, energy-efficient, flexible, scalable, and autonomous workpiece carriers described in Sect. 2.3 serve as a link between the individual production stations. Through the holistic detection of the environment, e.g. via ceiling cameras or low-cost integrated sensor systems, a cost-optimized design of the overall system can be ensured. The target price is approximately the cost of one meter of a double-strap conveyer belt. Through autonomous path planning and communication with the rest of the instances, the workpieces and machines of this system form the backbone of a cyber-physical material flow system (CPMS).

Establishing an Industry 4.0-compatible communication standard using enabling technologies described in Sect. 2.2 forms the basis of a central requirement specification of the production strategy. This framework is unaffected by details dispatched to specific jobs on specified machines at predetermined times. Using a high-level online controlling tool, production-related indicators in terms of cost, time, quality, resources, and energy can be visualized and evaluated. Combined with the simultaneous use of discrete event simulation for forward-controlling, substantial production KPIs as throughput times, utilization, delivery and punctuality, energy and resource needs, costs and results can be forecasted. Upgrading communications, material handling, and factory control in terms of cyber-physical systems is a prerequisite for the creation of a resilient factory, in which a production line is not linked to a product. It is thereby possible to flexibly adjust the processing stations to a changing product mix and capacity, thus optimizing overall utilization. Therefore, a new level of Jidoka can be realized, where the automation led by human as well as by machine intelligence is possible.

References

1. Bauernhansl T, ten Hompel M, Vogel-Heuser B (eds) (2014) Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung, Technologien, Migration. Springer Vieweg, Wiesbaden
2. Birkhofer R (2006) Einsatz von web-Technologien in der Automation. ZVEI, Frankfurt am Main
3. Birkicht A, Wilde J, Albrecht H, Detert M, Kowalski G, Pfeffer M, Schmidt B (2013) Technologie Roadmap Stressarme MST-Peckages: Trends, Perspektiven, Herausforderungen
4. Cnyrim H, Feuerstein R, Holzmann R (2014) Montagesysteme in der Elektronik. In: Feldmann K, Schöppner V, Spur G (eds) Handbuch fügen, Handhaben, Montieren, 2nd edn. Hanser, München, pp 770–798
5. Damm M, Leitner S-H, Mahnke W (2009) OPC unified architecture. Springer, Heidelberg
6. Eversheim W (1996) Organisation in der Produktionstechnik: Grundlagen. Studium und Praxis, Band 1, 3rd edn. VDI, Düsseldorf
7. Feldmann K (ed) (2009) Montage in der Leistungselektronik für globale Märkte: Design, Konzepte, Strategien. VDI-Buch. Springer, Berlin
8. Feldmann K (2014) Übersicht zu Montagesystemen. In: Feldmann K, Schöppner V, Spur G (eds) Handbuch Fügen, Handhaben, Montieren. Carl Hanser, München, pp 591–601
9. Feldmann K, Sturm J (1993) Yield improvement in SMT production by integrated process monitoring and testing. In: IEEE/CHMT (ed) Proceedings of the 15th IEEE/CHMT international electronics manufacturing technology symposium: "Electronics manufacturing for the year 2000". Institute of Electrical and Electronics Engineers, New York, pp 229–234
10. Feldmann K, Sturm J (1994) Closed loop quality control in printed circuit assembly. IEEE Trans Comp Hybrids Manuf Technol 17(2):270–276
11. Feldmann K, Schöppner V, Spur G (eds) (2014) Handbuch Fügen, Handhaben, Montieren. Handbuch der Fertigungstechnik. Carl Hanser, München
12. Franke J, Lütteke F (2012) Versatile autonomous transportation vehicle for highly flexible use in industrial applications. CIRP Ann Manuf Technol 61(1):407–410. doi: 10.1016/j.cirp.2012.03.048
13. Franke J, Risch F (2009) Effiziente Erstellung, Distribution und Rückmeldung von Werkerinformationen in der Montage. Z. wirtschaftlichen Fabrikbetrieb (ZWF) 104 (10):822–826
14. Franke J, Merhof J, Hopfensitz S (2010) Einsatz von dezentralen Multiagentensystemen Komplexitätsbeherrschung bei der Produktionsplanung und -steuerung. ZWF 105 (12):1075–1078
15. Goth C (2012) Analyse und Optimierung der Entwicklung und Zuverlässigkeit räumlicher Schaltungsträger (3D-MID)-Erlangen, vol 232. Meisenbach, Bamberg

16. Härter S, Klinger T, Beer D, Franke J (2016) Comprehensive correlation of inline inspection data for the evaluation of defects in heterogeneous electronic assemblies. In: SMTA (ed) Proceedings pan pacific microelectronics symposium, Kohala Coast, HI
17. Hedges M (2010) 3D aerosol jet printing—an emerging MID manufacturing process. In: 9th International Congress Molded Interconnect Devices, Fürth
18. Hon K, Li L, Hutchings IM (2008) Direct writing technology—advances and developments. CIRP Ann Manuf Technol 57(2):601–620. doi:10.1016/j.cirp.2008.09.006
19. Hörber J, Glasschröder J, Pfeffer M, Schilp J, Zäh M, Franke J (2014) Approaches for additive manufacturing of 3D electronic applications. In: Procedia CIRP: variety management in manufacturing—proceedings of the 47th CIRP conference on manufacturing systems, vol 17. Elsevier B.V., pp 806–811
20. Kagermann H, Wahlster W, Helbig J (2013) Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0. Promotorengruppe Kommunikation der Forschungsunion Wirtschaft–Wissenschaft. Frankfurt am Main
21. Liedl G (2014) Materialfluss zur Montage. In: Feldmann K, Schöppner V, Spur G (eds) Handbuch Fügen, Handhaben, Montieren, 2nd edn. Hanser, München, pp 420–432
22. Madisetti V, Bahga A (2014) Internet of things—a hands-on approach. VPT
23. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Hung Byers A (2011) Big data: the next frontier for innovation, competition, and productivity
24. Manyika J, Chui M, Bughin J, Dobbs R, Bisson P, Marrs A (2013) Disruptive technologies: advances that will transform life, business, and the global economy
25. Mayring P (2008) Qualitative Inhaltsanalyse. Beltz, Weinheim
26. Meinberg U, Topolewski F (1995) Lexikon der Fertigungsleittechnik: Begriffe, Erläuterungen, Beispiele. Springer, Berlin
27. Michl M (2013) Webbasierte Ansätze zur ganzheitlichen technischen Diagnose, Friedrich-Alexander-Universität, Lehrstuhl für Fertigungsautomatisierung und Produktionssystematik: Dissertation. Meisenbach, Erlangen-Nürnberg
28. Möller N (2008) Bestimmung der Wirtschaftlichkeit wandlungsfähiger Produktionssysteme. Forschungsberichte/IWB, vol 212. Utz, München
29. Moyne J, McAfee LC (1989) A Prototype SECS message service for communication in the semiconductor manufacturing environment. IEEE Trans Semicond Manuf
30. Müller J (2013) HTML5 programming with JavaScript for dummies. Wiley, New York
31. Pfeffer M, Reinhardt A (2009) Auswahlleitfaden für Montagelösungen in der Leistungselektronik. In: Feldmann K (ed) Montage in der Leistungselektronik für globale Märkte: Design, Konzepte, Strategien. Springer, Berlin, pp 83–170
32. Ramer C, Sessner J, Scholz M et al. (2015) Fusing low-cost sensor data for localization and mapping of automated guided vehicle fleets in indoor applications. In: IEEE (ed) Proceedings of the 2015 IEEE international conference on multisensor fusion and integration for intelligent systems (MFI), pp 65–70
33. Sauer W, Weigert G (2003) Modellierung von Fertigungsprozessen. In: Sauer W (ed) Prozesstechnologie der Elektronik: Modellierung, Simulation und Optimierung der Fertigung. Hanser, München, pp 21–77
34. Schindler G (2015) On the Smart Move: Industry 4.0 in Electronics Production. In: SMTA (ed) Proceedings surface mount technology association international conference (SMTAI), Rosemont, IL, pp 592–596
35. Schüßler F (2010) Verbindungs- und Systemtechnik für thermisch hochbeanspruchte und miniaturisierte elektronische Baugruppen. Fertigungstechnik-Erlangen, vol 216. Meisenbach, Bamberg
36. ten Hompel M, Schmidt T, Nagel L (2007) Materialflusssysteme: Förder- und Lagertechnik, 3rd edn. Springer, Berlin
37. Ullrich G (2012) Tragende Rolle im Triebwerksbau: MTU realisiert ein neues Montagekonzept für die Kleinserienmontage von Flugzeugtriebwerken. In: Logistik für Unternehmen (03/2012), pp 44–46

38. Westkämper E, Zahn E (eds) (2008) Wandlungsfähige Unternehmensstrukturen: Das Stuttgarter Unternehmensmodell. Springer, Berlin
39. Wolter K, Zerna T (2005) Produktionstechnik für eine Aufbau- und Verbindungstechnik für die Nanoelektronik. Fachverband Electronic Components and Systems im ZVEI—Zantralverband Elektrotechnik- und Elektronikindustrie e. V., Frankfurt am Main

Part II
Modeling for CPS and CMS

Cyber-Physical Systems Engineering for Manufacturing

Allison Barnard Feeney, Simon Frechette and Vijay Srinivasan

1 Introduction

Cyber-physical systems (CPS) are integrated embodiments of cyber systems (consisting of computing, communication, and control elements) and physical systems (consisting of geometrical and material elements). Almost all products used in modern society are cyber-physical systems. Almost all modern manufacturing systems to produce these products are also cyber-physical systems. Engineering such complex CPS has re-energized the field of systems engineering, which has been moving steadily away from a document-based practice to a model-based discipline. In fact, success in cyber-physical systems engineering strongly depends on proper application of model-based systems engineering (MBSE).

Manufacturing is a national priority in several countries, including the U.S.A. These countries are investing heavily in public-private partnerships in what they consider to be strategic, manufacturing-technology areas. The U.S. National Institute of Standards and Technology (NIST) is deeply involved in several *smart manufacturing systems* research projects that address standards and measurement-science problems in manufacturing systems. In these projects NIST is also applying advances in cyber-physical systems engineering to the manufacturing domain.

This chapter describes the critical link between modern manufacturing and cyber-physical systems engineering. The rest of the chapter is organized as follows.

A. Barnard Feeney · S. Frechette · V. Srinivasan (✉)
Engineering Laboratory, Systems Integration Division,
National Institute of Standards and Technology,
100 Bureau Drive, Gaithersburg, MD 20899, USA
e-mail: Vijay.Srinivasan@nist.gov

A. Barnard Feeney
e-mail: Allison.BarnardFeeney@nist.gov

S. Frechette
e-mail: Simon.Frechette@nist.gov

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series
in Wireless Technology, DOI 10.1007/978-3-319-42559-7_4

Section 2 describes the characteristics of cyber-physical systems. Section 3 deals with the evolution of systems engineering as a discipline to meet the challenges of engineering complex systems. Section 4 is concerned with the rise of manufacturing as a national priority, and the recent national efforts to boost research and development to enable manufacturing innovation. Section 5 describes the NIST projects in engineering smart manufacturing systems. Section 6 summarizes the chapter and offers some concluding remarks.

2 Cyber-Physical Systems

Cyber-physical systems are an inevitable consequence of the information revolution. Embedded computing, internet communication, and digital control have now become integral parts of modern engineered products and their manufacturing processes. Such products and processes are cyber-physical systems. The U.S. National Science Foundation (NSF) has been a major investor in fundamental research in CPS since 2010, and it defines and explains CPS as follows: "Cyber-physical systems are engineered systems that are built from, and depend upon, the seamless integration of computational algorithms and physical components. Advances in CPS will enable capability, adaptability, scalability, resiliency, safety, security, and usability that will far exceed the simple embedded systems of today. CPS technology will transform the way people interact with engineered systems—just as the Internet has transformed the way people interact with information. New smart CPS will drive innovation and competition in sectors such as agriculture, energy, transportation, building design and automation, healthcare, and manufacturing" [70].

Echoing this optimistic vision, a CPS Public Working Group (PWG) that was established by NIST says: "Cyber-physical systems are smart systems that include co-engineered interacting networks of physical and computational components. These highly interconnected systems provide new functionalities to improve quality of life and enable technological advances in critical areas, such as personalized health care, emergency response, traffic flow management, smart manufacturing, defense and homeland security, and energy supply and use" [63]. In mid-2014, NIST established the above mentioned CPS PWG to bring together a broad range of CPS experts in an open, public forum to help define and shape key characteristics of CPS. One of the goals of the group is to better manage the development and implementation of CPS within and across multiple *smart* application domains, including manufacturing, transportation, energy, and healthcare [11].

A couple of themes of interest to this chapter emerged from the NSF and the NIST-inspired descriptions of CPS. The first is that the CPS are networked systems. This immediately relates CPS to other major initiatives such as the Internet of Things (IoT) and the Industrial Internet [10]. The second theme is that the CPS are closely linked to advanced manufacturing. In fact, the German-led Industrie 4.0 initiative [42] predicts that the fourth industrial revolution will be based on CPS. A recent article in Harvard Business Review observes that "the United States stands to lead and benefit disproportionately in a smart, connected products world, given America's strengths in the core underlying technologies, many of the skills required, and key supporting industries. If this new wave of technology allows the U.S. to reinvigorate its capacity as a technology leader in the global economy, it will breathe new life into the American dream while contributing to a better world" [78].

3 Systems Engineering

While major technical advances are sweeping across CPS, the field of systems engineering is experiencing a renaissance. The International Council on Systems Engineering (INCOSE) defines and describes systems engineering as follows: "Systems Engineering is an interdisciplinary approach and means to enable the realization of successful systems. It focuses on defining customer needs and required functionality early in the development cycle, documenting requirements, then proceeding with design synthesis and system validation while considering the complete problem. Systems Engineering integrates all the disciplines and specialty groups into a team effort forming a structured development process that proceeds from concept to production to operation. Systems Engineering considers both the business and the technical needs of all customers with the goal of providing a quality product that meets the user needs." The renaissance in systems engineering referred to above is caused by the application of recent advances in information science and technology to the field of systems engineering.

Traditionally, the systems-engineering practice has been dominated by paper (or paper-equivalent electronic files) documents. Such documents are read only by human beings, who comprehend the content and take further action. This practice is being replaced by a systems-engineering discipline that is based on information models that can be read by machines (in addition to being read by humans). Machine-readability is a prerequisite for automation, which improves both the quality and speed of information processing in systems engineering.

In 2007 INCOSE published an influential document, called Systems Engineering Vision 2020, which outlined a vision for MBSE [31]. This vision heralded a transition from the prevailing document-based practice to a mode-based practice. This transition was aided by the development and deployment of open standards and software tools such as the Systems Modeling Language (SysML) for systems modeling and Modelica for systems simulation. The timing of these developments and tools is particularly opportune because of the significant increases in the complexity of CPS. These increases require formal modeling and simulation tools to define and analyze the systems with as much automation as possible—in other words, MBSE. The US-based Aerospace Industry Association recently outlined the benefits of leveraging MBSE across the entire product lifecycle including any government-industry collaborations for early requirements development. The U.S. Department of Defense is a major sponsor of a Systems Engineering Research Center to drive development and deployment of MBSE.

Recently, INCOSE has published an updated vision called Vision 2025 [32]. The context and content of Vision 2025 can be summarized as follows:

- The center of gravity for systems engineering has shifted from aerospace and defense sectors towards automotive and consumer electronics sectors. Such a trend is not surprising because automotive and consumer electronics companies have swiftly embraced CPS and have been investing heavily in research and development efforts in CPS.
- There is a move from MBSE to Model-based Enterprise (MBE) to cover life-cycle phases of products. Models created in the early phases of a product's life span using MBSE tools and principles are only of limited value if these models are not linked to those in design, manufacturing, testing, installation, service, and disposal phases of a product/system. MBE takes a much broader view of models and their interactions throughout a system's lifecycle.
- The focus of systems engineering has shifted to composition and integration, as opposed to mainly decomposition. The reason for this shift is the fact that industry and government are reluctant to undertake costly and time consuming projects that start with a clean sheet. The current emphasis is more on integrating and testing existing subsystems and technologies in rapid iterative cycles. The notion of system decomposition is still important, but it is no longer the primary driver. This trend is also in line with the *spiral* (rapid build-and-test iterations) development process instead of the *V-shaped* (top-down decomposition followed by bottom-up composition) development process.

There is a general acknowledgment among the practitioners that systems engineering is still largely a powerful book-keeping exercise. But, it is being transformed into a *smart* book-keeping exercise throughout the manufacturing sector.

4 Manufacturing Innovation

In the United States, it is widely acknowledged that the U.S. government laboratories and universities conduct world-class research, while industry focuses on product development and commercialization. This leaves an increasing gap in applied research for manufacturing, which deals primarily with moving advanced manufacturing technologies from research to production. Without this *missing middle*, good ideas can often get lost. This recognition led the United States President's Council of Advisors on Science and Technology (PCAST) to identify the following transformative technologies for manufacturing innovation: advanced sensing and control; informatics and visualization; digital manufacturing integration; and advanced materials manufacturing [76, 77].

In response to the PCAST recommendations to spur manufacturing innovation, industry and government organizations have formed new programs to address smart manufacturing technology and infrastructure development. The Smart Manufacturing

Leadership Coalition [84] has developed a smart manufacturing platform architecture, and has outlined technology and standards priorities for smart manufacturing. SMLC's technology priorities include modeling and simulation, sensor integration, data collection and management, and enterprise systems integration. The Industrial Internet Consortium [30] is focused on enabling smart technologies for industrial applications. The IIC is working to accelerate the growth of the Industrial Internet by promoting best practices, fostering the creation of industry test beds, and developing reference architectures and frameworks necessary for interoperability. The U.S. National Network for Manufacturing Innovation [69] has put in place several manufacturing institutes including the Digital Manufacturing and Design Innovation Institute [14]. Digital manufacturing is the ability to connect different parts of the manufacturing life-cycle through data, and to utilize that information to make smarter, more efficient business decisions.

Such manufacturing innovation initiatives are not restricted to the United States. In Europe, the goal of the Germany's Industrie 4.0 project is to develop the intelligent factory (Smart Factory), characterized by adaptability, resource efficiency, and ergonomics. Its technological bases are cyber-physical systems and the IoT. In Japan, the Ministry of Economy, Trade and Industry (METI) announced a strategy for *Smart Convergence* to develop technology-independent and *Leading-Edge Integrated Industries* through digitization and networking.

It is in this context of world-wide interest in manufacturing innovation that NIST has embraced *smart manufacturing* as having the potential to contribute to the public good by fundamentally changing how products are designed, manufactured, used, and retired. Smart manufacturing systems will produce less waste, use less energy, consume fewer resources, and provide more business opportunity. The rapid adoption of internet connectivity, wireless technologies, cloud-based storage, and data analytics has already stimulated the growth of smart manufacturing systems. To realize the full potential of smart manufacturing, new technologies and new standards are needed.

5 Smart Manufacturing Systems Programs at NIST

The NIST Smart Manufacturing programs align with the NIST mission to promote U.S. innovation and industrial competitiveness. Currently there are two smart manufacturing systems programs at NIST—one focusing on the *design and analysis* problems, and the other on the *operations planning and control* problems. Systems thinking and cyber-physical systems engineering are pervasive throughout these programs. These programs address key opportunities identified by the PCAST for dramatically rethinking the manufacturing process with advanced technologies and shared infrastructure. In these programs, NIST is bridging the gap between industry requirements and fundamental measurement science through delivery of standards and reference data derived from NIST research and strong industry collaborations.

The value of standards to industry and the economy is underscored in the PCAST report [77]. Standards "spur the adoption of new technologies, products and manufacturing methods. Standards allow a more dynamic and competitive marketplace, without hampering the opportunity to differentiate. Development of standards reduces the risks for enterprises developing solutions and for those implementing them, accelerating adoption of new manufactured products and manufacturing methods."

Standards are a critical tool for leveling the playing field for small businesses by reducing the cost barriers. Product Lifecycle Management (PLM) standards, for example, contribute to both agility (by streamlining processes) and quality (by enabling the integration of different activities along the product and production-system lifecycles). Standardized interfaces make open source and low cost PLM systems possible. In the production-systems area, device-connectivity standards are enabling small businesses to provide machine performance and systems-reliability solutions to improve productivity, quality, and sustainability. Standards for enterprise and supply-chain systems integration, such as the Open Applications Group Integration Specification (OAGIS), help streamline business processes between partners in the supply chain. These standards enable low cost applications that are appropriate for small manufacturers to work with enterprise-level applications used by Original Equipment Manufacturers (OEM).

Different standards contribute in different ways to realize smart manufacturing systems. There are a vast number of standards in this space, and a critical survey of these standards, their adoption and applicability is something that industry is seeking. A NIST-developed standards landscape [50, 51] takes a first step. The landscape defines key, smart-manufacturing capabilities and presents a smart-manufacturing ecosystem. The smart-manufacturing ecosystem encompasses three dimensions—products, production systems, and enterprise (business) systems. The landscape associates standards with the lifecycle phases of all three dimensions. We will now describe the two major smart manufacturing systems programs at NIST.

5.1 Smart Manufacturing Systems Design and Analysis

The research program for Smart Manufacturing Systems Design and Analysis (SMSDA) is organized into the following four projects of investigation: (1) modeling methodologies for manufacturing system analysis, (2) predictive analytics, (3) performance measurement for manufacturing systems, and (4) service-based manufacturing and service composition.

Based on previous NIST work on sustainable manufacturing, unit process modeling [54, 55], and manufacturing services [82] the SMSDA program seeks to develop an analytical framework for design, analysis, and prediction of manufacturing systems. The program focus is primarily discrete part manufacturing and assembly, but does include batch-type manufacturing applications. The program is developing formal methods and tools for dynamic composition of manufacturing component models to facilitate prediction and performance measurement. To predict and measure production-system-level performance for different manufacturing scenarios, unit models need to be dynamically compositional in an analytic environment that represents the larger production system. In addition, system architectures that enable computer-processible, service-description models are being investigated for their utility in integration of manufacturing systems. Such models are necessary to facilitate automated registration, discovery, and composition of manufacturing services within agile manufacturing systems. Methods of verification, validation, and uncertainty quantification for these models are also being studied. The four projects in the SMSDA program are described in the next four subsections.

5.1.1 Modeling Methodologies for Manufacturing System Analysis

When developing new, and operating existing, production systems, manufacturers require knowledge that the proposed system designs are feasible and will yield optimal results. Rather than using analytical models, many manufacturers still use empirical (e.g., trial and error) methods to design, operate, or redesign production systems. There are several reasons. First, the development of models and the interpretation of results do not follow a precise methodology shared across various usages. This limits the ability to develop systematic means to apply analytical techniques to decision making. It also significantly increases the time and cost of making actionable recommendations. Because of this, analytical modeling efforts for manufacturing systems are often overtaken by events such as decision deadlines and equipment malfunctions, among others.

Second, in addition to creating analytical models, there are fundamental challenges to actually using them. Challenges include (1) making use of information from various sources, (2) knowing that the techniques to be applied are appropriate to the situation, (3) knowing the extent to which analytical results are valid, and (4) acting on the insight the effort provided. The consequences include missed opportunities to reuse knowledge, unreliable results, and high cost of analysis.

Currently, the models and information sources used in analytical activities are not easily integrated. A principal barrier to integration is a lack of methods that support composition both among model components and disparate viewpoints [12]. A viewpoint is a set of related concerns drawn from a representation of the whole system. Existing research does not take advantage of the unique characteristics of smart manufacturing. These characteristics include on-going need for analytical methods, their integration with data from operations, their integration with production control systems, and the ability to dedicate a portion of manufacturing resources to experimental investigation of new processes.

The variety of problems to which analytical methods may be applied in manufacturing makes specifying widely-applicable integration strategies difficult.

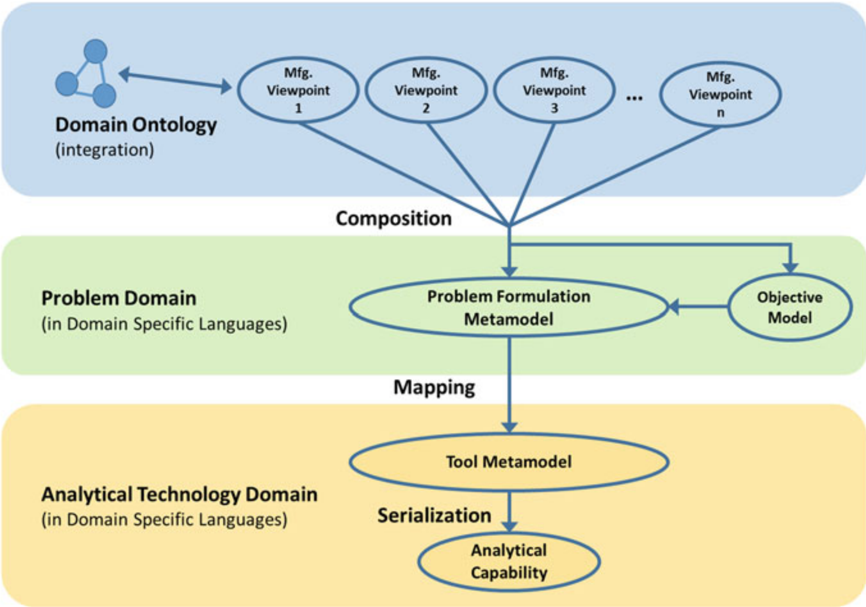


Fig. 1 Developing analytical capability by composing domain-specific viewpoints

An emerging technology, domain-specific modeling, provides an efficient means to represent a variety of viewpoints, but it lacks a methodology for effective composition of model components. Figure 1 illustrates how such models could be composed to form sophisticated analytical tools by using problem-formulation and tool meta-models. Equation-based modeling offers an effective method of composition, but it only works for certain applications.

One of those applications, and the initial focal point of this project, is optimization, specifically scheduling optimization. The project is developing methods for representing and composing the analytical models needed to formulate and solve scheduling problems in a smart-manufacturing plant [13]. The methods involve synthesizing elements of functional, domain-specific, and equation-based modeling methods. If successful, this methodology, which can be extended to deal with other types of optimization problems, will become an integral part of smart manufacturing systems.

5.1.2 Predictive Analytics for Manufacturing Systems

The dramatic increase in the availability of data from the machine to the enterprise has increased the potential for improved prognostics and diagnostics. Predictive analytics is the principal foundation for realizing that potential. Numerous

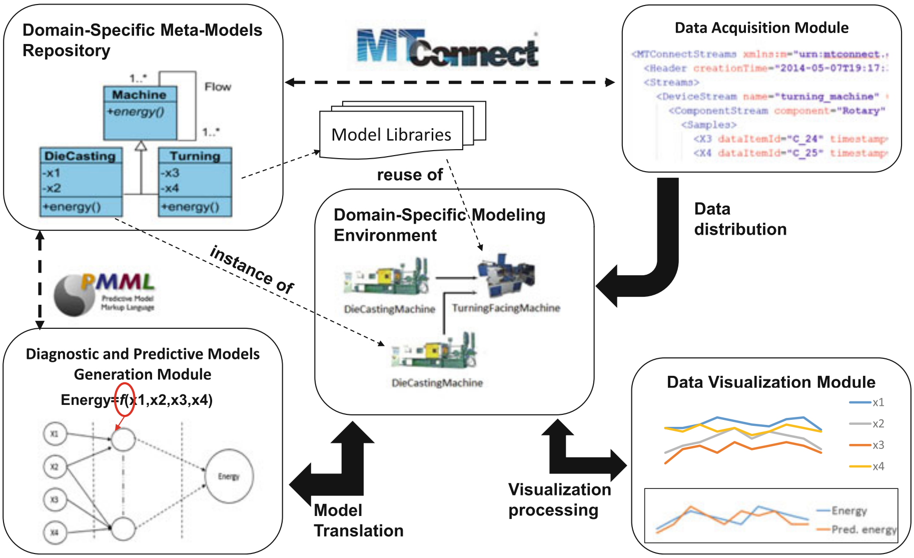


Fig. 2 Overview of a framework for predictive analytics in manufacturing

commercial, predictive-analytics solutions are in use today. These existing solutions, however, are based on proprietary models that run on open source platforms. These solutions are limited to large OEMs and are rarely available as reconfigurable, open applications suited for small and medium enterprises.

This project seeks to remove this limitation by developing open protocols and standards for the data inputs to those solutions and measurement methods for characterizing and evaluating their results. Specifically, the project will develop (1) solutions for data capturing, fusion, dimension reduction, and filtering and (2) a measurement-based approach to address challenges of traceability, uncertainty quantification, security, verification, validation, and data provenance. The goal of the project is to use those methods and standards to construct the predictive capabilities needed for both prognostics and diagnostics from the machine-level to the enterprise-level.

Standards for analytics information, such as Predictive Markup Modeling Language (PMML), are being extended to support manufacturing applications and are a major focus of this project [46]. Figure 2 shows an overview of a proposed framework for predictive analytics in manufacturing [45]. The vision of this project is to use this framework as a foundation for demonstrating a prototype, predictive-analytics solution that improves production system efficiency. The prototype system will include manufacturing models for predictive analytics, domain-specific languages for performing predictive analytics, and standard interfaces for data analytics tools. The goal of this prototype is to make it easier for manufacturing domain experts to build manufacturing system models and generate the necessary analytical models from a manufacturing system specification.

5.1.3 Performance Measurement for Smart Manufacturing

Manufacturers are adopting smart systems to drive gains in agility, productivity, quality, and sustainability. These smart systems integrate information and communication technologies with intelligent software applications to optimize a variety of performance metrics that result in the on-time delivery of customized, high-quality products. Being able to determine effective metrics and measure actual performance, therefore, is critical to achieving that result.

Two important questions arise. The first is, "What metrics are used to characterize a given *performance*?" The second is, "How does one quantify that metric using realizable measurements?" The original, and long-standing, performance criteria was productivity. Then came quality—thanks in large part to the success of the Japanese in the 1980s. In the era of smart manufacturing, two more criteria have been added to the list. Agility is related to how fast and how well manufacturers adapt to changes in the market. The most recent addition, and the focus of much of this research, is sustainability [43].

Sustainability in manufacturing is defined as the creation of manufactured products through processes that are non-polluting, conserve energy and natural resources, and are economically sound and safe for employees, communities, and consumers. Metrics for sustainability are not as mature as other metrics and this is an active area of research [38]. As productivity and agility of manufacturing systems increases, the necessity for better understanding and controlling the sustainability-related impacts of those systems increases. Manufacturing sustainability may be expressed in terms of environmental impact with primary focus on the efficient use of energy and natural resources [79]. Sustainable manufacturing is a challenging problem because reductions in resource consumption and environmental impacts must balanced against other drivers including timeliness, quality, productivity, and cost. Understanding the changes to a system in terms of multiple objectives is made more difficult when criteria against which sustainability assessments can be made are neither measured nor available in such a way as to be shared at a system level [54].

This project focuses on developing standard metrics and measurement methods for enabling smart design and analysis of production systems. A fundamental challenge faced by the manufacturers is identifying opportunities for improving performance and integrating new smart technology to realize those improvements. Industry routinely collects operational-level data of all kinds. What industry lacks is the knowledge needed to use that data to improve overall performance [44]. The goal of this project is to develop the knowledge needed to measure and standardize practices for evaluating manufacturing performance. Sustainability is given particular attention as the least-understood driver for smart manufacturing. In pursuit of that goal, reference architectures, standard representation methods, and *crowd-sourced* knowledge collection solutions are being explored.

In the early stages of the project, two activities dominated. The first involved an investigation into the methods for identifying operational improvements within manufacturing systems in a series of papers on the topic [27, 28, 39, 40, 41]. The results of that investigation led to a study of how to use models of the manufacturing system, referred to as *reference architectures* to understand performance assurance. Earlier NIST work [2] was used as a basis for illustrating opportunities for improvement [40]. This approach was then extended to a new reference architecture for factory design and improvement [9]. This model will be a foundation for future standards and guidelines for promoting more effective factory design and improvement practices.

In the second activity, NIST led the development of two guides: ASTM E60.13–Standard Guide for Evaluation of Sustainability of Manufacturing Processes, to standardize methods for evaluating the performance of manufacturing processes, and ASTM E60.13–Guide for Characterizing Environmental Aspects of Manufacturing Processes, to standardize methods for characterizing the performance of manufacturing processes as building blocks for system analysis with specific focus on sustainability evaluation [1]. Prior work on terminology for sustainable manufacturing [22, 62, 61] is fundamental to measuring performance and may be a key contribution to the ASTM standard on terminology for sustainable manufacturing.

These two activities are still ongoing with an emphasis on the use and extension of the standards [56, 80, 84]. In addition, under NIST's leadership, ASTM has initiated a new work item titled Standard Guide for the Definition, Selection, and Composition of Key Performance Indicators to Evaluate Environmental Aspects of Manufacturing Processes. Currently the project is exploring two approaches for collecting and disseminating a broad base of manufacturing knowledge to industry. The first approach involves the creation of a national repository of unit-manufacturing-process models based on the ASTM E60 guidelines [3]. The second is to cull the implementation knowledge of practitioners to understand where, when, and how to apply smart manufacturing technologies in the field [28]. Documenting implementation knowledge of this type has been shown to successfully move the state of the art forward and to help vendors identify and remedy recurring problems in the cybersecurity domain [58].

5.1.4 Service-Based Manufacturing and Service Composition

Traditional manufacturing systems spanned the lifecycle of a product by the integration of software subsystems through a combination of open and proprietary exchange of data. The rapid revolution and adoption of industrial internet technologies is replacing software subsystems with manufacturing-based software services. Cloud computing is the principal technology enabling this revolutionary change.

Cloud computing is enabling an eco-system of composable (easy to assemble and reassemble) manufacturing services that will accelerate new product development, gain efficiencies in production and supply chain management, and allow use of data analytics to optimize manufacturing activities. This changes the traditional integration paradigm substantially. In a recent workshop [64] industry participants discussed the nature of these changes, the technical challenges of addressing them, and their potential benefits. The overwhelming conclusion of the participants was that the major benefit of open, dynamically composable, cloud services will be a new standards-based platform that will advance innovations in smart manufacturing systems.

One of the technical challenges discussed extensively at the workshop, and the focus of this project, is service discovery. Participants argued that before services can be composed, they must be discovered. They agreed that discovery involves two steps. The first involves specifying service requirements and service capabilities that allow representation and registration. The second involves developing metrics and algorithms that can match one to the other. This project's initial efforts center on the first step. The project is developing reference models, analysis methods, and synthesis tools as a basis for standards-related specifications of both requirements and capabilities. The results will reduce the risks to service providers, cloud vendors, and manufacturing users by providing tools, based on the reference architectures, that guide the development and validation of such standards.

This research is pursuing a computational, model-driven approach for specifying manufacturing services requirements and capabilities. This approach is believed to enable generation of computer-processable representations that will facilitate efficient registration, discovery, and, eventually, composition of services [37]. An example of such a tool, based on the ISO Standard Core Component Specification (CCS) meta-model [35], is the NIST Messaging Standard Semantic Refinement Tool (MSSRT). Figure 3 shows the application of the MSSRT within the OAGIS standard to aid the service providers and users in generating and cataloging the messaging standard usage information using a new, CCS-compliant OAGIS meta-model. The usage information includes a human readable implementation guideline as well as machine readable message exchange specifications in various formats. The figure also illustrates development of an OAGIS reference business process meta-model and Business Process Cataloging and Classification System (BPCCS) to allow automated generation of both standard and context-specific OAGIS business processes. The reference business process meta-model is extending ideas found in earlier industrial initiatives such as ebXML [21, 35]. These concepts can be applied to many similar standards.

To develop the envisioned computer-processable representations, the project advances some of industry-driven context-specification and management approaches. Differing proposals for context handling and/or context-based document configuration have been reviewed, such as the ones found in the ebXML specifications [89]. In one case, an approach was focused on proposing context meta-model and defining possible contextual categories and appropriate classifications that can be used to define the list of allowed values for each context

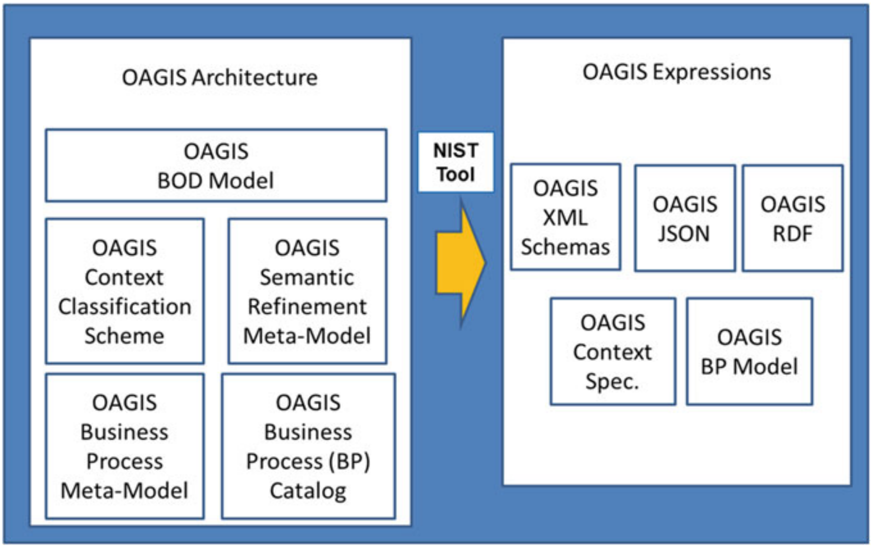


Fig. 3 NIST Semantic Refinement and Business Process Modeling Tool used to automatically generate alternate implementation models from meta-model of OAGIS Standard

category. This specification may be further improved, for example, by expanding context meta-model to support different association types between classification values. In the other case, an approach defines a method for business document assembly and context-specific derivation of the assembled document. Expanding the previous approaches, the essential, new ideas in our approach are 1) to use contextual information to enable formal, precise cataloging and life-cycle management of the messaging specification usage information and 2) to develop a system that enables sharing and use of such information to drive the evolution of the core messaging standard. These ideas are being implemented using the MSSRT and BPCCS tools.

5.2 *Smart Manufacturing Operations Planning and Control*

The Smart Manufacturing Operations Planning and Control (SMOPAC) Program focuses standards and measurement science to support integration and technology challenges in the factory. Within the SMOPAC program, smart manufacturing systems are defined to be fully-integrated, collaborative manufacturing systems that respond in real time to changing customer demands and operating conditions in the factory. The success of smart manufacturing systems depends upon the ability to easily and rapidly reconfigure factory production and supply networks to optimize system performance. Such systems must deal effectively with uncertainty and abnormal events and learn from past experience to enable continuous improvement. These systems must enable seamless interoperability between small, medium, and large manufacturers. The complexity of the overall challenge is due to:

- Complex system, sub-system, and component interactions within smart manufacturing systems make it challenging to determine specific influences of each on process output metrics and data integrity.
- Lack of uniform processes that guide manufacturing operations management, integrated wireless technologies, prognostics and diagnostics, and cybersecurity at all levels (from component to system). Many existing solutions are currently proprietary and are seldom disseminated.
- Simultaneous operations of systems increase the intricacy and understanding of information flow relationships

The SMOPAC research plan consists of a portfolio of interrelated projects that focus on key research areas: (1) digital thread, (2) systems analysis integration, (3) wireless systems, (4) cybersecurity, and (5) prognostics, health management and control. In addition to the five projects, a Smart Manufacturing Systems (SMS) Test Bed provides an integrated testing environment and a source of real manufacturing data for internal and external researchers. The program has developed a conceptual framework for lifecycle information management and the integration of emerging and existing technologies, which together form the basis of our research agenda for dynamic information modeling in support of digital-data curation and reuse in manufacturing [24]. Collectively these activities provide a comprehensive approach that leads to new industry standards and practices. The five projects of the SMOPAC program and the SMS Test Bed are described in the following six sections.

5.2.1 Digital Thread for Smart Manufacturing

The promise of smart manufacturing cannot be achieved without access to the right data at the right time. Today's industrial practice lacks visibility of product and process information across lifecycle functions. Design functions lack visibility into information about manufacturing processes that will be used to produce the product. Manufacturing functions lack visibility into the intent of the design engineer or measurement results of early production runs. The Digital Thread for Smart Manufacturing project is concerned with (1) making semantically-rich product and process data available through open standards, (2) establishing data quality, certification and traceability, and 3) using trusted information to build knowledge and enable better decision making. This project builds upon past NIST work in MBE and MBSE and is the largest of the five SMOPAC projects.

Figure 4 illustrates the opportunities for information sharing and integration in the portion of the lifecycle that is the focus of this project. Standards exist that support the integration of product and process information for systems that implement similar functions. Examples include all Computer-Aided Design

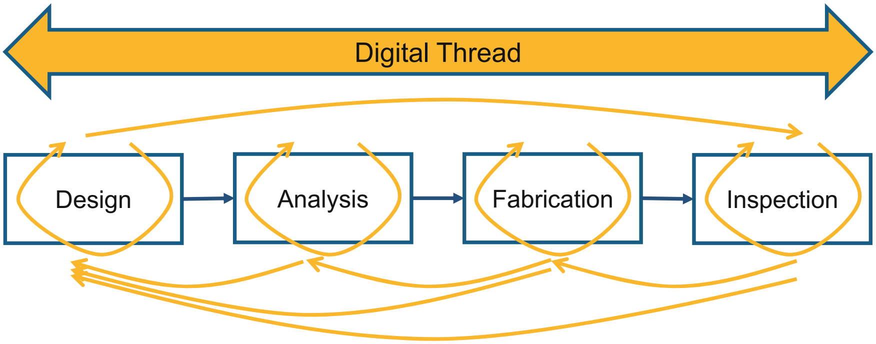


Fig. 4 Information flows enabled by digital manufacturing

(CAD) systems, all computer-aided manufacturing (CAM) systems, all computer-aided engineering (CAE) systems, and all product data management (PDM) systems. The most widely recognized of these integration standards is called STEP [34]. Another such standard is the Quality Information Framework (QIF) [15]. This project additionally seeks to support integration of heterogeneous systems that make up the entire product lifecycle, using open standards, enabling industry to move from model-based design to MBE. This project is currently analyzing workflows to learn what data is common amongst different models in different stages of the lifecycle. These common elements support interfaces between systems. This project calls the integration of heterogeneous information models and these additional common elements the *common information model* [81].

NIST plays a significant role in the development of QIF and STEP standards. NIST led the US effort to support the transition from model-based design to model-based enterprise by leading the development of the latest STEP application protocol ISO 10303-242:2014 (AP242) [36], whose capabilities are illustrated in Fig. 5. AP242 contains computable representations for several types of 3-dimensional (3D) model data, including geometric dimensioning and tolerancing (GD&T) information [18]. This information conveys the design intent and functional requirements of the product to manufacturing. The intent is for AP242 to support all product and manufacturing information (PMI) needed by manufacturing and inspection planning activities. NIST is guiding the development of a second edition of AP242 that will add new representations for electrical wire harness, kinematics, and additional PMI.

NIST has provided both the leadership and significant technical contribution to the development of the QIF standard. The American National Standards Institute (ANSI) recently approved a new edition, QIF v2.0 as an standard (DMSC 2015b). This new edition enhances the previous edition by providing a complete and accurate 3D product definition with (1) semantic geometric and dimensional tolerances, (2) definitions for measurement resources, (3) a template for measurement rules, and (4) statistical functionality [57, 59]. These new capabilities satisfy

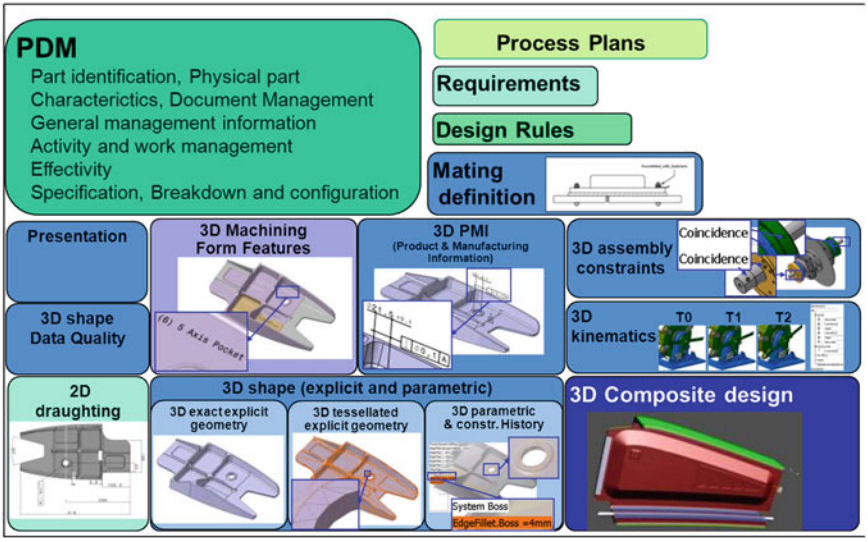


Fig. 5 Capabilities of STEP AP242

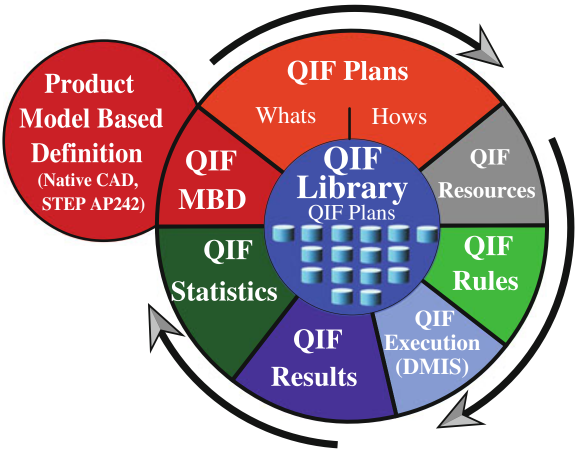


Fig. 6 Quality information framework

the digital interoperability needs for a wide variety of smart-manufacturing use cases including (1) feature-based dimensional metrology, (2) quality measurement planning, (3) first article inspection, and (4) discrete quality measurement. The quality management system at most modern factories includes operations that based on, what are called today, cyber-physical systems. These advanced technological systems are required to gather and use digital manufacturing data constantly. QIF defines, organizes, and associates that data with into higher–level information objects. Such objects (see Fig. 6) include measurement plans, results, part geometry, PMI, measurement templates, resources, and statistical analysis [15].

Pilot projects are part of the project's strategy to increase adoption of open standards and smooth the road to MBE. One pilot discovered much of the barrier to full adoption of MBE is cultural. Despite being in the digital age, a large percentage of small- and medium-sized suppliers still receive OEM designs as full-detail, 2-dimensional (2D) drawings or as a combination of 3D-shape-geometry models plus 2D drawings containing the PMI [23]. A large percentage of suppliers must either remodel the part completely or add the PMI manually to the imported shape-geometry model. Much of the CAD industry has implemented STEP AP242 with embedded PMI, reducing the need for drawings. The same degree of implementation has not occurred in the CAM and Coordinate Measuring System (CMS) industries. NIST conducted a pilot project that demonstrated the value of improved CAD-to-CAM and CAD-to-CMS data interoperability using STEP AP242 with embedded PMI. This pilot provided seed funding to software solution providers to develop interfaces to open standards. Additional results from this pilot include identified gaps in open standards and metrics on time savings through MBE [19, 88].

The second focus of this project is to ensure data quality and build trust in data across the lifecycle. Ensuring data quality is critical to the digital enterprise. In addition to participating in the development of AP242, NIST has developed a strategy for measuring conformance of CAD systems to the American Society of Mechanical Engineers suite of standards for PMI [20, 48]. NIST developed sets of PMI test-case models now being used by both industry and solution providers to develop internal processes, determine recommended practices, and test conformance of software applications [49]. NIST also developed software for analyzing STEP physical files [47] that are used by STEP solution providers. Currently, the project is investigating the use of embedded digital certificates in standard data formats for authentication, authorization, and traceability of product data. While digital certificates are used widely in other domains, they are not widely used in engineering and manufacturing. Extensions to common standard formats STEP, QIF, and MTConnect [60], have been developed and are in different stages of approval in the different standards communities. A digital manufacturing certificate toolkit has is available as open-source software [66]. Use of embedded digital certificates in PLM workflows will increase trust in product data throughout the lifecycle.

The third focus of this project is using lifecycle information to make decisions. Once information is collected and stored in a semantically-rich and computer-interpretable manner, and its quality and provenance are known, the information can be acted upon for better decision making. For example, the design development activity is often supported by a design knowledge base. The knowledge base contains meta-data, rules, standards, correlations, and design dictionaries. Industry lacks a way to discover data relationships and link data across the lifecycle. Such a data observatory would enable near-real-time dynamic updating of domain-specific knowledge bases using machine learning and artificial intelligence methods. This project is defining a conceptual framework of emerging and existing technologies that can make lifecycle information available and actionable.

5.2.2 Systems Analysis Integration for Smart Manufacturing Operations

The Systems Analysis Integration for Smart Manufacturing Operations project seeks to deliver methods and protocols for (1) unifying discipline-specific, engineering, analysis information and (2) integrating it with existing, unified, systems-modeling information that is modeled in a formal modeling language. SysML [75] is a such a language. Moreover, it is a standard and is also widely used around the world. This project uses higher-level system models, created in SysML, to coordinate discipline-specific engineering analysis. Coordination is achieved by identifying and eliminating inconsistencies between the system-level models and analysis-level models. The goal of the project is to enable systems modeling tools and discipline-specific analysis tools to efficiently exchange and use information during smart manufacturing operations.

NIST has a long history of involvement in the development of a variety of formal modeling languages in the Object Management Group (OMG). In particular, we have focused on the continued evolution of the SysML. Most recently, NIST led development of information models for system-operation requirements, product-family variation modeling, and computer interchange of graphical representations. NIST also provided software to assess models and modeling-tool compliance to SysML and related standards [4, 71–75]. These standards are the basis of the integrations undertaken in this project.

Systems engineering models contain system requirements, designs, and tests, often specified in graphical modeling languages, such as SysML. These models must be developed in conjunction with analysis models, such as those used to simulate both physical interactions and numeric signal flows by solving a set of differential equations. System-engineering and simulation models are typically developed in separate modeling tools, reducing the efficiency of engineering processes. This project integrated both the physical-interaction and signal-flow simulation modeling into SysML. To do so, the project reviewed physical-interaction and signal-flow simulation tools to develop a common abstraction of their constructs and semantics. This abstraction was compared to SysML and missing simulation concepts were identified. The project team also proposed an extension to SysML to address the gaps, and gave examples of their application [5]. These extensions will be brought to the OMG for standardization. In subsequent work, we will develop logical models for finite element simulation and integrate those models with system models.

5.2.3 Wireless Systems for Industrial Environments

The Wireless Systems for Industrial Environment project develops integrated methodology and protocols to enable, assess, and assure the real-time performance of wireless systems in industrial environments. An industrial environment, such as one found in a smart manufacturing environment, requires a variety of wireless

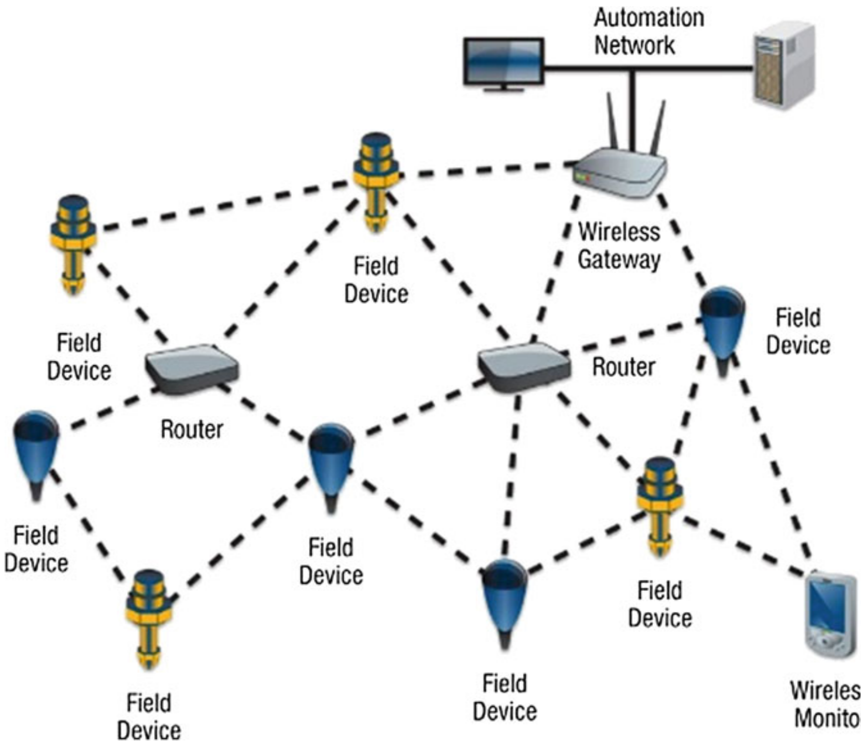


Fig. 7 A wireless sensor network

technologies to provide seamless connectivity from low-power sensor nodes to high data rate video links. This project focuses on standards-based wireless protocols used in industrial environments. However, the metrics, methodology, and guidelines developed are applicable to proprietary wireless protocols as well.

Wireless sensor networks (WSN) are a key technology for IoT in manufacturing. As shown in Fig. 7, a WSN is an internet-like network of sensor nodes that cooperatively sense and possibly control the environment autonomously or with people in the loop. The advent of smaller, cheaper, rugged and low-powered sensors is bringing the IoT to even the smallest objects. NIST, with the International Electrotechnical Commission (IEC), held a workshop to identify challenges of WSNs for IoT. The results of the workshop are laid out in a white paper published by the IEC [29]. The paper details the fragmented and disjoint standards landscape for wireless networks and stresses the need for increased communication and coordination among different standard organizations, unified planning, optimized resource allocation and reduced repetition of work.

The equipment at each level of a factory equipment network (factory level, work-cell level, and device level) may employ different network communication protocols. It is a challenge for industry to efficiently and effectively test all kinds of equipment with different network communication protocols at different network levels. To address this challenge, the project developed a test bed for designing and evaluating performance metrics of operational systems that are instrumented with standard and non-standard wireless technologies. A key contribution of the wireless testbed is methodology by which wireless performance is correlated with operational performance for circumstances where wireless technologies are used for supervisory control or feedback control [7]. We refer to this as bridging Information Technology (IT) performance to Operational Technology (OT) performance.

Real-time sensor data from WSN is essential for making decisions in controlling industrial processes and condition monitoring. However, wireless communication is subject to interference and thus may affect critical industrial operations. The project has developed simulation framework in a wireless test bed to study how various wireless sensor network configurations and topologies affect the performance, including safety, of manufacturing plant operations. The first test case was simulation of a continuous process chemical plant operation where sensor output is interfaced to an IEEE 802.15.4-based wireless sensor network via a programmable logic controller. The integration of a simulated physical system with a real wireless network provides the ability to examine the effects of real-time wireless communications in a factory running different wireless activities on simulated plant processes [6].

In addition to understanding how wireless technologies work, the project is helping manufactures make better decisions regarding those technologies. Field measurements and channel models are being used to optimize not only selections but also their actual configuration in factories. Using modern state-of-the-art RF (Radio Frequency) sounding techniques, the project has measured RF propagation in several real machine shops—the one at NIST, several in partner manufacturing facilities, and one in a contract machine shop. Interference levels were measured using precision, spectrum analyzers in an effort to correlate interference with the happenings on the plant floor. RF propagation was measured using high-precision, time-synchronized, RF-sounding equipment developed by NIST engineers. The sounding equipment collects raw propagation information required to compute statistical channel models of factory environments. High volumes of raw data are being post-processed to produce complex-valued correlations (impulse responses) that represent the black-box propagation losses and distortions of each sounding scan.

The RF measurement data will be interpreted using standard channel models with the precise parameterization required to accurately characterize RF propagation in the factory. The channel models, raw impulse response data, and packet error rate curves provide a basis for detailed network simulations that are integrated with models of physical processes. These integrated simulations will allow NIST to study the impacts of the RF environment and wireless networking technologies on the performance of physical plant processes.

5.2.4 Cybersecurity for Smart Manufacturing Systems

The Cybersecurity for Smart Manufacturing Systems project seeks to determine quantitatively the impact of cybersecurity on real-time performance, resource use, reliability, and safety of smart manufacturing systems. This project focuses on two research challenges: (1) the development of comprehensive requirements and use cases that represent practical cybersecurity approaches for real-world needs, and (2) the development of a suite of specific tests that measure the impact of cybersecurity technology when fulfilling these needs.

In its initial stages, the project conducted a two-day Roadmapping Workshop on Measurement of Security Technology Performance Impacts for Industrial Control Systems (ICS) at NIST. The 66 participants represented a balanced cross-section of ICS stakeholder groups, including manufacturers, technology providers, solution providers, university researchers, and government agencies. The workshop report [16] serves as a foundation for the development of a measurement-science research for ICS security at NIST. This project is developing new methods and metrics for measuring the performance impact of security technologies [86].

Specifically, project team members are working with standards development organizations to develop new guidelines and standards to facilitate the implementation of cybersecurity technologies that do not negatively impact performance—see NIST Special Publication 800-82 [87]. This major report includes guidance on: (1) recommendations on ICS risk management, (2) how to tailor traditional IT security controls to accommodate unique ICS performance, reliability, and safety requirements, (3) on threats and vulnerabilities, (4) recommended practices, (5) security architectures, and security capabilities and tools. The guidance will enable improved ICS security in manufacturing and critical infrastructure industries, while simultaneously addressing the demanding performance, reliability, and safety requirements of these systems.

The project has developed a smart-manufacturing-system cybersecurity test bed to implement test methods that analyze both network and operational performance impacts of proposed cybersecurity safeguards and countermeasures. Those methods are in accordance with the best practices and requirements prescribed by national and international standards and guidelines such as NIST Special Publication 800-82 [87]. Figure 8 shows the physical design of the cybersecurity test bed that includes a chemical reactor process and a robotic assembly process, which are two key areas in manufacturing [8]. In the next phase of the test bed deployment, a transportation enclave will be installed.

Cybersecurity communities have created a variety of data representation and exchange standards. These standards address weaknesses and vulnerabilities, naming conventions, system state, configuration checklists, asset identification, and severity measurement of software and configuration systems. NIST has developed an infrastructure, the Security Content Automation Protocol [65], for leveraging this array of information standards. This infrastructure provides technical guidance for how the existing standards should be used together; however, there are gaps in the

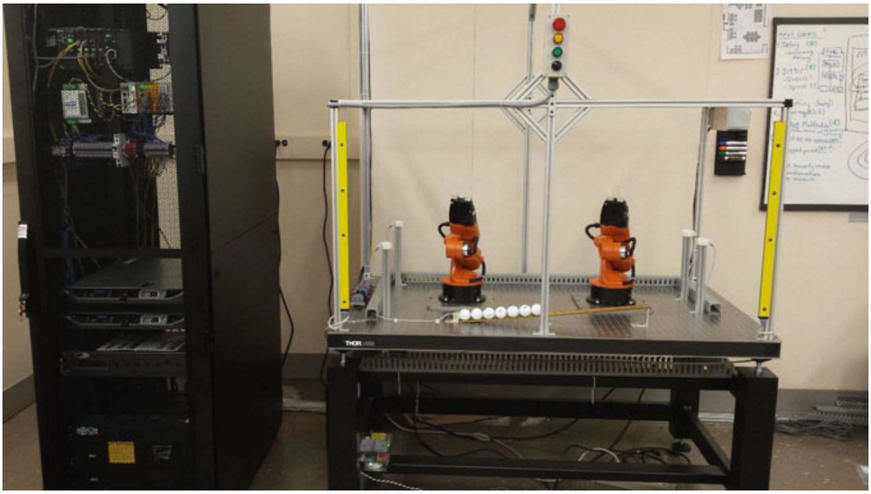


Fig. 8 A fully assembled discrete manufacturing robotics enclave

standards. [52] developed a user interface for selecting and tailoring security controls in accordance with NIST SP 800-53.

One future plan is to extend this user interface to be in accordance with NIST SP 800-82 security requirements to trace security control selections to the NIST Cybersecurity Framework Core functions and outcomes. Another plan is to develop a manufacturing implementation (Profile) of the NIST Cybersecurity Framework (CSF). The Manufacturing CSF Profile will be implemented in the smart manufacturing system cybersecurity test bed using various cybersecurity solutions to measure any network and operational performance impacts. From this research, guidance will be developed on implementing the CSF in manufacturing environments without having negative performance impacts on the systems.

5.2.5 Prognostics, Health Management and Control

The goal of the Prognostics, Health Management (PHM), and Control project is to develop methods, protocols, and tools for robust sensing, diagnostics, prognostics, and control. These results will enable manufacturers to respond to planned and unplanned performance changes, which will enhance the efficiency of smart manufacturing systems. Early implementations of smart manufacturing technologies enable manufacturers to use equipment and process data to inform decision-makers to determine the impact on both performance and overall process health and update their maintenance strategies. There is an increasing interest to leverage this data in concert with data from emerging sensing technologies to generate diagnostic and prognostic intelligence for improved control. This project focuses on the standards and measurement science needed to enable and promote such intelligence.

Complex system, sub-system, and component interactions within smart manufacturing systems make it challenging to determine the specific influences of each on process performance, especially during disruptions. The simultaneous operation of complex systems within the factory increases the difficulty to determine and resolve failures due to ill- or undefined information flow relationships. There is no uniform process that guides sensing, PHM and control at all levels. Proprietary solutions exist that integrate some manufacturing systems, but they apply to systems from select vendors and are often expensive and inaccessible to smaller manufacturers. The goal of this project is to promote advanced sensing, PHM, and control from ISA 95 [33] manufacturing levels 0 (production process) through 3 (manufacturing operations management). This will result in improved decision-making support and greater automation with a focus on vendor-neutral approaches and plug-and-play solutions.

At the outset of the project, a review was conducted of PHM-related standards to determine the industries and needs addressed by such standards, the extent of these standards, and any similarities as well as potential gaps among the documents. The results of that assessment can be found in [90]. The project then conducted a workshop to elicit the needs and priorities of stakeholders in the PHM technology arena. The attendees identified and prioritized measurement science needs for improving PHM impacts within manufacturing processes; measurement science barriers, challenges, and gaps that prevent the broad use of PHM technologies for manufacturing processes; and the research and development needed to address the priority measurement and standards challenges. The workshop resulted in a report that highlights roadmaps that will advance the state-of-the-art in manufacturing PHM [17]. Key findings from the workshop and report include critical measurement science challenges and corresponding roadmap [92].

Prior to the workshop, three critical research thrust areas were identified: machine tool linear axes diagnostics and prognostics, manufacturing process and equipment monitoring, and PHM for robotics. The workshop findings, and resultant roadmap, reinforced the necessity of these research thrusts and strengthened our specific approach with targeted, first-hand knowledge from the manufacturing community. These areas are briefly described below.

- The machine tool linear axes diagnostics and prognostics research thrust focuses on developing a sensor-based method to quickly estimate the degradation of linear axes without disrupting production [91]. The method, which has been demonstrated in a newly-constructed linear-axis test bed at NIST (1) detects translational and angular changes due to axis degradation, (2) supports verification and validation of similar PHM techniques and (3) produces reference data sets that can be used by PHM developers as test data. These data sets are valuable to manufactures so that they may test their systems without risking damaging to or impacting the productivity. This method will ultimately lead to standards that monitor the health of and predict degradation linear axes. Developing diagnostics and prognostics of linear axes enables optimization of maintenance scheduling and part quality.

- The manufacturing process and equipment monitoring research thrust focuses on identifying high-value data sources for systems-level PHM and developing methods to collect this data to avoid the challenges associated with big data. The idea is to have the right data at the right time for analysis and control. This research is supported by a systems-level test bed (described later in the chapter) of networked machine tools and sensors in an active manufacturing facility. Early efforts have concentrated on integrating sensors and machine tool controllers with production management systems using data exchange standards, such as MTConnect.
- The PHM for robotics research thrust focuses on developing methods, metrics, assessment protocols, and reference data sets for industrial robot arm systems. These resources will enable manufacturers to (1) detect robot-system-performance degradation and (2) predict how such degradation impacts key elements of the robot system (e.g., accuracy). A PHM-focused robotics test bed, including industrial robotic arms, is being developed to support this research thrust. This work concentrates on the health of the arm and the overall robotic system (e.g., arm, controller, sensors, end-effector) as well as the relationships between the performance of these components and the performance of the system.

This research is being supported by external collaborations where appropriate. For example, have developed a hierarchical methodology that will enable manufacturers to appropriately decompose their complex manufacturing systems into individual components for monitoring and maintenance. They use a systematic approach that enables a manufacturer to identify and assess the risks of fault and failure of their components, processes or systems [53].

5.2.6 Smart Manufacturing Systems Test Bed

To support the five projects mentioned above, NIST has undertaken an effort to develop a Smart Manufacturing Systems (SMS) Test Bed [25, 68] as illustrated in Fig. 9. The SMS test bed is a test bed for the program. This SMS test bed provides a *smart manufacturing system* as a way to test the results of the projects all together, all at once. Creating such a test bed, however, requires access to real industrial systems to understand the real problems with collecting, transmitting, analyzing, and acting on data and information quickly and reliably throughout the entire smart manufacturing system.

Prior manufacturing test beds including [26, 83] focused primarily on production and ignored the larger product lifecycle. The goal of this test bed is to extend the production-focused test bed concepts in the past to include testing the other phases of the product lifecycle. To achieve that goal, the test bed includes a Computer-Aided Technologies (CAx) Lab, a Manufacturing Lab and data publication web services. These labs are integrated using a string of digital interfaces creating a *digital thread* of information across the product lifecycle. This test bed serves as a reference implementation that manufacturers may use to collect data safely and

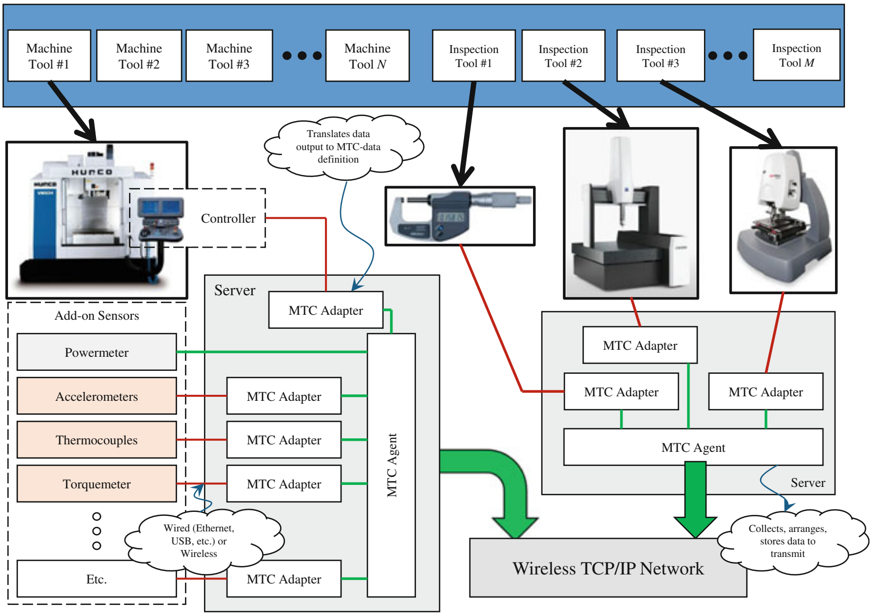


Fig. 9 Architecture of NIST smart manufacturing systems test bed

efficiently without disruption to operations. Data is collected from the Manufacturing Lab using the MTConnect standard. That data is aggregated and published internally and externally of NIST via web services. Three channels of data dissemination are available or becoming available from the SMS Test Bed: (1) a volatile data stream using an MTConnect agent, (2) a query-able data repository using the NIST Material Data Curation System [67], and (3) pre-compiled data packages that include a collection of CAx Lab data and associated Manufacturing Lab data.

6 Summary and Concluding Remarks

In this chapter we described the convergence of cyber-physical systems, systems engineering, and manufacturing innovation, and how NIST is responding to address the standards and measurement science issues caused by this convergence. While much progress is being made in these areas, there are some emerging trends that will require further research and development. These trends include:

- Smart requirements engineering. Eliciting, analyzing, and communicating requirements of complex manufacturing systems continue to be a major challenge. Performing these tasks quickly and correctly will contribute considerably to the success of manufacturing innovation initiatives.

- Dynamics of interacting multi-domain systems. We need more research into mathematical modeling of multiple domains and analysis of their dynamic interactions. This requires people with different subject matter expertise to collaborate, and create new knowledge and tools.
- Visualization and integration of humans. Explosion of data in manufacturing requires smart visualization tools that will enable humans to be properly integrated with complex systems.
- Affordable solution for small- and medium-sized enterprises. Reducing the cost of ownership and training in complex engineering information systems is emerging as a major challenge even in technologically advanced countries.
- Checking and testing composability and compositionality. We need more and better tools for automating the process of checking and testing complex systems as they are composed and integrated into larger systems.

Certain commercial systems are identified in this chapter. Such identification does not imply recommendation or endorsement by NIST; nor does it imply that the products identified are necessarily the best available for the purpose.

References

1. ASTM (2016) ASTM E60.13 sustainable manufacturing. http://www.astm.org/commit/subcommit/E6013.htm
2. Barkmeyer E, Christopher N, Feng S, Fowler J, Frechette S, Jones A, Scott H (1996) SIMA reference architecture, part 1: activity models. National Institute of Standards and Technology, NISTIR 5939
3. Bernstein W, Mani M, Lyons K, Morris K (2016) An Open web-based repository for capturing manufacturing process information. In: ASME 2016 international design engineering technical conferences and computers and information in engineering conference
4. Bock C, Barbau R, Narayanan A (2014) BPMN profile for operational requirements. J Object Technol 13(2):2:1–2:35
5. Bock C, Matei I, Barbau R (2015) Integrating physical interaction and signal flow simulation with systems engineering models. NAFEMS World Congress, San Diego, CA
6. Candell R, Lee K (2015) Measuring the effect of wireless sensor network communications on industrial process performance. In: 2015 ISA process control and safety symposium, Houston, TX
7. Candell R, Lee K, Proctor F (2015a) Software-defined radio based measurement platform for wireless networks. In: International ieee symposium on precision clock synchronization for measurement, control and communication 2015, Beijing, China
8. Candell R, Zimmerman T, Stouffer K (2015b) An industrial control system cybersecurity performance testbed. National Institute of Standards and Technology. NISTIR 8089
9. Choi S, Jung K, Kulvatunyou S, Morris K (2016) A survey of existing technologies and standards: design and operation of a smart manufacturing system. J Res Natl Inst Stan
10. Chui M, Löffler M, Roberts R (2010) The internet of things. McKinsey Quarterly
11. CPS (2015) Public Working Group. http://www.cpspwg.org/
12. Denno P, Blackburn M (2015) Using semantic web technologies for integrating domain specific modeling and analytical tools. Procedia Comput Sci 61:141–146
13. Denno P, Kim D (2015) Integrating views of properties in models of unit manufacturing processes. Int J Comput Integr Manuf. doi:10.1080/0951192X.2015.1130259

14. DMDII (2016) Digital manufacturing and design innovation institute. http://dmdii.uilabs.org/
15. DMSC (2016) Quality information framework. http://qifstandards.org/
16. Energetics (2014) Roadmap for measurement of security technology impacts for industrial control systems. Energetics, Inc., Gaithersburg, MD
17. Energetics (2015) Measurement science roadmap for prognostics and health management for smart manufacturing system. Energetics, Inc., Gaithersburg, MD
18. Barnard Feeney A, Frechette S, Srinivasan V (2015) A portrait of an ISO STEP tolerancing standard as an enabler of smart manufacturing systems. J Comput Inf Sci Eng 15(2):021001
19. Fischer K, Rosche P, Trainer A (2015) Investigating the impact of standards-based interoperability for design to manufacturing and quality in the supply chain. NIST Government Contractor Report (NIST GCR), National Institute of Standards and Technology
20. Frechette S, Jones A, Fischer B (2013) Strategy for testing conformance to geometric dimensioning tolerancing standards. In: 12th CIRP conference on computer aided tolerancing, Huddersfield. Elsevier, United Kingdom
21. Fuger S, Najmi F, Stojanovic N (2005) ebXMLl registry information model version 3.0. http://docs.oasis-open.org/regrep/v3.0/specs/regrep-rim-3.0-os.pdf
22. Garretson I, Mani M, Leong S, Lyons K, Haapala K (2016) Manufacturing process characterization terminology to support sustainability assessment. J Cleaner Prod
23. Hedberg T Jr, Krima S, Camelio J (2016) Embedding X.509 digital certificates in three-dimensional models for authentication, authorization, and traceability of product data. In: ASME 2016 international design engineering technical conferences and computers and information in engineering conference
24. Hedberg T Jr, Barnard Feeney A, Helu M, Camelio J (2016) Towards a lifecycle information framework and technology in manufacturing. J Comput Inf Sci Eng
25. Helu M, Hedberg T (2015) Enabling smart manufacturing research and development using a product lifecycle test bed. Procedia Manuf 1:86–97
26. Helu M, Robinson S, Bhinge R, Banziger T, Dornfeld D (2014) Development of a machine tool platform to support data mining and statistical modeling of machining processes. In: Machine tool technologies research foundation 2014 annual meeting, San Francisco, CA
27. Helu M, Morris K, Jung K, Lyons K, Leong S (2015) Identifying performance assurance challenges for smart manufacturing. Manuf Lett 6:1–4
28. Helu M, Libes D, Lubell J, Lyons K, Morris K (2016) Enabling smart manufacturing technologies for decision-making support. In: ASME 2016 international design engineering technical conferences and computers and information in engineering conference
29. IEC (2014) Internet of things: wireless sensor networks. International electrotechnical commission, Geneva, Switzerland
30. IIC (2016) Industrial internet consortium. http://www.iiconsortium.org/
31. INCOSE (2007) Systems engineering vision 2020, INCOSE-TP-2004-004-02
32. INCOSE (2014) Systems engineering vision 2025: A world in motion
33. ISA (2016) ISA95, Enterprise-control system integration. https://www.isa.org/isa95/
34. ISO (1994) ISO 10303 Industrial automation systems and integration—product data representation and exchange—part 1: Overview and fundamental principles. International Organization for Standardization, Geneva, Switzerland
35. ISO (2014) ISO 15000 electronic business extensible markup language (ebXML)—part 5: core components specification. International Organization for Standardization, Geneva, Switzerland
36. ISO (2014) ISO 10303 Industrial automation systems and integration—product data representation and exchange—part 242: application protocol: managed model-based 3D engineering. International Organization for Standardization, Geneva, Switzerland
37. Ivezic N, Kulvatunyou B, Srinivasan V (2014) On architecting and composing through-life engineering information services to enable smart manufacturing. In: Proceedings of the 3rd international conference in through-life engineering services, Procedia CIRP, vol 22, pp 45–52
38. Joung C, Carrella J, Sarkar P, Feng S (2013) Categorization of indicators for sustainable manufacturing. Ecol Ind 24:148–157

39. Jung K, Morris K, Lyons K, Leong S, Cho H (2015) Mapping strategic goals and operational performance metrics for smart manufacturing systems. Procedia Comput Sci 44:184–193
40. Jung K, Morris K, Lyons K, Leong S, Cho H (2015) Using formal methods to scope performance challenges for smart manufacturing systems: focus on agility. Concurr Eng 2 (4):343–354
41. Jung K, Morris K, Lyons K, Leong S, Cho H (2016) Performance challenges identification method for smart manufacturing systems. National Institute of Standards and Technology, NISTIR 8108
42. Kagermann H, Wahlster H, Helbig J (2013) Recommendations for implementing the strategic initiative industrie 4.0. Industrie 4.0 Working Group
43. Kibira D, Morris K, Kumaraguru S (2015) Methods and tools for performance assurance of smart manufacturing systems. National Institute of Standards and Technology, NISTIR 8099
44. Kumaraguru S, Kulvatunyou S, Morris K (2014) Integrating real-time analytics and continuous performance management in smart manufacturing systems. In: Advances in production management systems. Innovative and knowledge-based production management in a global-local world. Springer, Heidelberg, pp 175–182
45. Lechevalier D, Narayanan A, Rachuri S (2014) Towards a domain-specific framework for predictive analytics in manufacturing. In: IEEE international conference on big data, issue, 27–30 October 2014
46. Li Y, Roy U, Lee Y-T, Rachuri S (2015) Integrating rule-based systems and data analytics tools using open standard PMML. In: ASME 2015 international design engineering technical conferences and computers and information in engineering conference
47. Lipman R (2014) STEP file analyzer user's guide (version 2). National Institute of Standards and Technology, NISTIR 7999
48. Lipman R, Lubell J (2015) Conformance checking of PMI representation in CAD model STEP data exchange files. Comput Aided Des 66:14–23
49. Lipman R, Lubell J, Frechette S (2014) MBE PMI validation and conformance testing. http://www.nist.gov/el/msid/infotest/mbe-pmi-validation.cfm
50. Lu Y, Morris K, Frechette S (2015) Standards landscape and directions for smart manufacturing systems. In: 2015 IEEE international conference on automation science and engineering
51. Lu Y, Morris K, Frechette S (2016) Current standards landscape for smart manufacturing systems. National Institute of Standards and Technology, NISTIR 8107
52. Lubell J (2015) Extending the cybersecurity digital thread with XForms. In: Balisage: the markup conference 2015, Washington, DC, vol 15
53. Malinowski M, Beling P, LaViers A, Marvel J, Weiss B (2015) System interdependency modeling in the design of prognostic and health management systems in smart manufacturing. In: Annual conference of the prognostics and health management society 2015, Coronado, CA
54. Mani M, Johansson B, Lyons K, Sriram R, Ameta G (2012) Simulation and analysis for sustainable product development. Int J Life Cycle Assess
55. Mani M, Madan J, Lee J, Lyons K, Gupta S (2014) Sustainability characterization for manufacturing processes. Int J Prod Res 52(20):5895–5912
56. Mani M, Larborn J, Johansson B, Lyons K, Morris K (2016) Standard representations for sustainability characterization of industrial processes. J Manuf Sci Eng
57. Michaloski J, Hedberg T Jr, Huang H-M, Kramer T (2016) End-to-end quality information framework (QIF) technology survey. National Institute of Standards and Technology, NISTIR 8127
58. MITRE (2016) CWE—Common Weakness Enumeration. https://cwe.mitre.org/
59. Morse E, Heysiattalab S, Barnard Feeney A, Hedberg T Jr (2016) Interoperability: linking design and tolerancing with metrology. In: 14th CIRP conference on computer aided tolerancing, Gothenburg, Sweden
60. MTConnect (2016) MTConnect. http://www.mtconnect.org/

61. Narayanan A, Lechevalier D, Morris K, Rachuri S (2013) A methodology for handling standards terminology for stustainable manufacturing. National Institute of Standards and Technology, NISTIR 7965
62. Narayanan A, Lechevalier D, Morris K, Rachuri S (2015) Communicating standards through structured terminology. Comput Stand Interfaces 40:34–41
63. NIST (2015a) CPS framework. http://www.cpspwg.org/Portals/3/docs/CPS%20PWG%20Draft%20Framework%20for%20Cyber-Physical%20Systems%20Release%200.8%20September%202015.pdf
64. NIST (2015b) NIST workshop on open cloud architectures for smart manufacturing. http://www.nist.gov/el/msid/opencloudarch.cfm
65. NIST (2015c) Security content automation protocol. http://scap.nist.gov/revision/index.html
66. NIST (2016a) Digital manufacturing certificate toolkit. https://github.com/usnistgov/DT4SM/tree/master/DMC-Toolkit
67. NIST (2016b) NIST materials genome initiative. https://mgi.nist.gov/materials-data-curation-system
68. NIST (2016c) Smart manufacturing systems test bed. http://www.nist.gov/el/facilities_instruments/smstestbed.cfm
69. NNMI (2013) National network for manufacturing innovation: a preliminary design
70. NSF (2015) Cyber-physical systems. http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503286
71. OMG (2011) Business Process Notation and Model 2.0, Object Management Group
72. OMG (2014a) Diagram Definition 1.1, Object Management Group
73. OMG (2014b) UML Profile for BPMN Processes, Object Management Group
74. OMG (2015a) Unified Modeling Language 2.5, Object Management Group
75. OMG (2015b) Systems Modeling Language 1.4, Object Management Group
76. PCAST (2012) The President's Council of Advisors on Science and Technology, Capturing Domestic Competitive Advantage in Advanced Manufacturing AMP Steering Committee Report
77. PCAST (2014) President's Council of Advisors on Science and Technology, Report to the President on Accelerating U.S. Advanced Manufacturing: AMP 2.0 Steering Committee Report
78. Potter M, Heppelmann J (2014) How smart, connected products are transforming competition. Harvard Business Review, November 2014
79. Rachuri S, Morris K (2013) Sustainable manufacturing program workshop report, National Institute of Standards and Technology, NISTIR 7975
80. Rebouillat L, Barletta I, Johansson B, Mani M, Bernstein W, Morris K, Lyons K (2016) Understanding sustainability data through unit manufacturing process representations: a case study on stone production. In: 49th CIRP conference on manufacturing systems
81. Ruemler S, Zimmerman K, Hartman N, Hedberg T, Barnard Feeney A (2016) Promoting model-based definition to establish a common information model. J Manuf Sci Eng
82. Shin J, Kulvatunyou B, Lee Y, Ivezic N (2013) Concept analysis to enrich manufacturing service capability models. Procedia Comput Sci 16:648–657
83. Simpson J, Hocken R, Albus J (1982) The automated manufacturing research facility of the national bureau of standards. J Manuf Syst 1(1):17–32
84. SMLC (2016) Smart manufacturing leadership coalition. https://smartmanufacturingcoalition.org/
85. Smullin M, Haapala K, Mani M, Morris K (2016) Using industry focus groups and literature review to identify challenges in sustainable assessment theory and practice. In: ASME 2016 international design engineering technical conferences and computers and information in engineering conference
86. Stouffer K, Candell R (2014) Measuring impact of cybersecurity on the performance of industrial control systems. Mech Eng 136(12):59–62

87. Stouffer K, Pillitteri V, Lightman S, Abrams M, Hahn A (2015) Guide to industrial control systems (ICS) security. National Institute of Standards and Technology, NISTSP 800(82), Rev. 2
88. Trainer A, Hedberg T, Barnard Feeney A, Fischer K, Rosche P (2016) Gaps analysis of integrating product design, manufacturing, and quality data in the supply chain using model-based defintion. In: 2016 ASME manufacturing science and engineering conference, Blacksburg, VA
89. UN (2000) CEFACT: Electronic business XML (ebXML) technical architecture specification. Draft v 0.65
90. Vogl G, Weiss B, Donmez A (2014) Standards for prognostics and health management (PHM) techniques within manufacturing operations. In: 2014 annual conference of the prognostics and health management society, Fort Worth, TX
91. Vogl G, Weiss B, Donmez A (2015) A sensor-based method for diagnostics of machine tool linear axes. In: 2015 annual conference of the prognostics and health management society, Coronado, CA
92. Weiss B, Vogl G, Helu M, Qiao G, Pellegrino J, Justiniano M, Raghunathan A (2015) Measurement science for prognostics and health management for smart manufacturing systems: key findings from a roadmapping workshop. In: Annual conference of the prognostics and health management society 2015, Coronado, CA

Model-Based Engineering of Supervisory Controllers for Cyber-Physical Systems

Michel Reniers, Joanna van de Mortel-Fronczak and Koen Roelofs

1 Introduction

The notion of cybermanufacturing, as defined in [22], refers to future manufacturing systems that are expected to develop into complex, networked cyber-physical systems either in one physical location or distributed across many. The physical components of such systems, which can be mechanical, robotic, chemical or electrical, are going to be fully interoperable and driven by computer models of product data, systems, and processes. In this context, to enable the networked integration of manufacturing machines, equipment, and systems, coordination control is one of the important aspects. Model-based development of the coordination control layer has the potential to enable reconfigurability in strongly integrated and networked environments.

1.1 Model-Based Systems Engineering

The performance and functionality of complex integrated machines or cyber-physical systems depend on the strong interaction between the control system and the physical components. During the development process of this kind of systems this interaction should be taken into account from the beginning. The process starts with a thorough analysis of user requirements that describe the required functionality and performance. From the user requirements, concept specifications for both the physical and the control components are derived during the concept development phase.

M. Reniers (✉) · J. van de Mortel-Fronczak · K. Roelofs
Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: m.a.reniers@tue.nl

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_5

Using the requirements as a starting point, the dynamic behavior of the physical components and a proper way to control them are specified. For this purpose, increasingly often models of the physical components and of the control system are developed, simulated, and analyzed [10]. The emphasis in this stage lies primarily on validating whether the user requirements are fulfilled. The specifications form the guidelines for the next phases of the development process. Therefore, it is very important that the specifications are correct, meaning that if the control system and the physical components each work according to its own specification, then the system should fulfil the user requirements. Simulation is a powerful tool that can be used in the concept-development phase to determine, describe, and validate specifications. As simulation can be used early in the development process, simulation-based validation reduces the risk of expensive concept-design errors.

The physical-design phase is meant to precisely define how the control system interacts with the sensors and actuators. Models with an accurate description of the sensors and actuators of the system are developed, and the control system is adapted to match the changes. Moreover, the model can be used for evaluation of system performance trade-offs. For example, the selection of controller hardware components, such as processors or bus structures, can be well motivated as control structure/complexity and communication intensity are known and validated in the concept model.

In the implementation and testing phase, the control system is moved to the real-time platform. For correct real-time functionality, the calculations should be performed within well-defined time periods. Therefore, it is important to assess the impact of the duration of calculations on communication behavior. In case of differences with respect to the behavior in simulation time, correctness analysis has to be performed, which can give rise to changes in the design. Eventually (after downloading) the control system operates the real machine through an I/O interface. The first two development phases deliver virtual system models, which are accurate descriptions of system components and of the associated control. Ideally, in the last phase, the virtual system model is replaced by its real counterpart, and the control system is applied in the real-time environment without modifications.

Depending on the purpose, models can be formulated on different abstraction levels. Additionally, if a proper abstraction level is chosen, the implementation of validated control models can be a straightforward process. Especially in the case of cyber-physical systems, it is important to determine a suitable form of models necessary to perform efficient simulation experiments or even to successfully synthesize required control. When modelling physical components, a certain abstraction has to be made of the physical behavior of the actuators and sensors present in the machine. Most important is that the control system does not differentiate between the virtual and the real physical components. A movement, for instance, can be modelled by events representing its start and completion giving rise to discrete-event component models. The same movement can also be modelled using differential equations giving rise to continuous-time component models. Because sensors are usually modelled as discrete-event components, in the latter case, the corresponding system model is hybrid.

In the first place, specifying and modelling of systems and their control is essential for the concept-design phase. However, since it became possible for properly defined control models to be applied in a real-time environment without modifications, system and control models are of importance in the implementation and testing stage, as well. Namely, they can be used in various forms of model-in-the-loop, hardware-in-the-loop or software-in-the-loop testing, as shown in [7, 15, 34]. For example:

- If hardware or its prototype are available, they can be tested in combination with the real-time implementation automatically generated from the controller model. This speeds up the test process, because the design of the controller, for instance, for the prototype test setup is fast and easy.
- In principle, the real-time implementation of the controller can be used to control the hardware (system) directly. However, often the controller should be implemented on a dedicated, low cost embedded microcontroller. Real-time testing of this embedded microcontroller, even if no hardware is available, is possible as the virtual hardware model can be used for this purpose.
- If only a part of the hardware is available, real-time models of the remaining hardware components and the controller can be used to test only the available hardware components. The model of the hardware components now simulates in real-time the communication interaction with the hardware component that should be tested. This allows an incremental hardware development approach, where first one component is designed, built and tested before the design of the next component is started. Of course, a similar approach can be followed for the controller development.

Since cyber-physical systems consist of many components performing their actions in parallel, it seems justifiable to use a specification formalism which exploits this parallel character: for instance, Petri nets [24], (hybrid) process algebras [3] or networks of (hybrid) automata [14]. The general idea is that systems are treated as collections of independent components that interact by synchronizing on time, by synchronizing events or by sharing variables. A recent survey presented in [18] provides a summary of modelling techniques and tools proposed for the representation and the design of cyber-physical systems and their architectures. Although the evidence shows that many relevant aspects are addressed individually (representation, specification, control synthesis, simulation, verification) and different suitable types of models are introduced, no unifying framework is reported.

In this chapter, we discuss the usage of the CIF modelling and simulation language and toolset developed at Eindhoven University of Technology (http://cif.se.wtb.tue.nl/), which are meant to support the cyber-physical system design workflow [33]. The CIF language allows the specification of networks of hybrid automata for modelling uncontrolled systems, requirements and supervisors. The toolset provides simulations (random and interactive) with visualization to support the validation of all types of models involved. The tool provides state-of-the-art synthesis algorithms and has model transformations to external tools that support verification. All activities of the workflow demonstrated in this chapter are performed with support of the CIF toolset. Other tools such as Matlab/Simulink and Modelica (see [13, 26]) may equally well be used for activities such as modelling and simulation of hybrid systems, but lack support for supervisory controller synthesis.

Other system engineering approaches exist in literature of which SysML [12] and MechatronicUML [4] are prominent examples. The main difference between these approaches and CIF is that CIF offers concrete support for synthesis and analysis of models whereas the mentioned systems engineering formalisms merely facilitate integration of heterogeneous models.

1.2 Structure of This Chapter

In Sect. 2, the cyber-physical system design workflow is elaborated with special focus on supervisory control and the synthesis procedure. The case study used for the illustration is shortly described in Sect. 3. In Sect. 4, hybrid automata models of physical components and their abstraction to discrete-event automata are described. Section 5 focuses on models of requirements. Supervisor synthesis is discussed in Sect. 6. In Sect. 7, simulation-based visualization and its role in the development process are considered. Concluding remarks are presented in Sect. 8.

2 Synthesis-Based Development of Coordination Control

The control of cyber-physical systems usually consists of several layers of controllers. As mentioned above, coordination control is one of the important aspects because it enables the networked integration of manufacturing machines, equipment, and systems. In this chapter, we focus on supervisory controllers at the coordination layer and the interface to feedback controllers at the resource layer. As mentioned in [8], the feedback control loops are based on continuous representations of components. At higher layers, discrete-event representations are suitable for dealing with situations like system or work cycle start-up and shut-down, task initiation and coordination, change of operation mode, exception handling, failure diagnosis and recovery. The intermediate layer, interface, is positioned between feedback controllers designed based on continuous-time models of the system and the control logic implemented by the supervisory controller. In this setting, the system and feedback controllers from the lower layers can be abstracted as discrete-event models for the purpose of supervisory control. At the interface, information from sensors and feedback controllers is abstracted in the form of events, while command events from the supervisory controller are translated to appropriate input signals to the actuators or set-points to feedback controllers. The position of the supervisory control layer in a cyber-physical (manufacturing) system is graphically depicted in Fig. 1, which is inspired by [6].

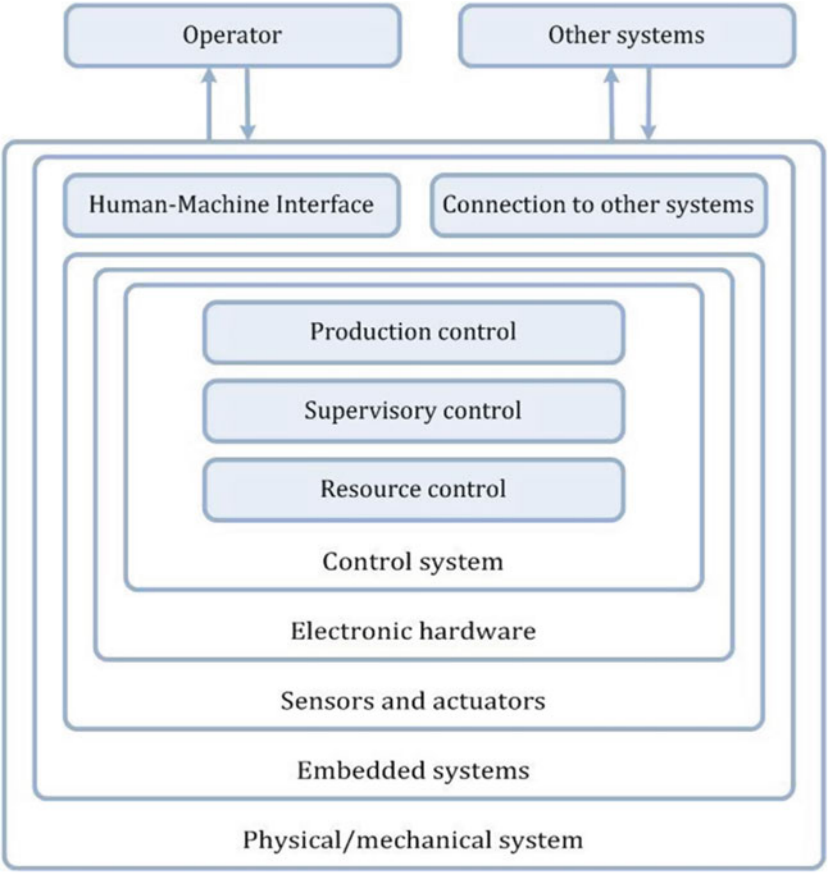


Fig. 1 Positioning of supervisory control

Above, we explained that models play an important role in the development process and that different kinds of models can be used for different purposes. An overview of types of models essential for supervisory control development, along with relevant process steps and model relationships, is given in Fig. 2.

Based on this overview, the following workflow is proposed.

1. To start with, hybrid automata models [14] of (physical) system components are developed, called uncontrolled hybrid plant. These models describe all possible behaviors the components can exhibit, not restricted for a specific system function. Simulation and simulation-based visualization (using an image model of the system) can be used to validate these models.
2. From the uncontrolled hybrid plant, the uncontrolled discrete-event plant can be abstracted (possibly with the help of a hybrid observer that abstracts information

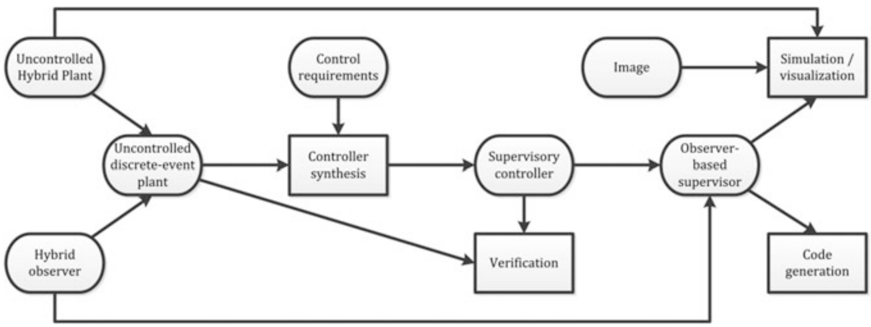


Fig. 2 Overview of models in the development process [33]. *Oval shapes* indicate models used in the development process and *rectangular shapes* indicate activities in the development process

from sensors and feedback controllers by events). These models are expressed in terms of extended finite automata [29].

3. Then models of requirements (related to the function the system should fulfil) are defined in terms of extended finite automata or state-based expressions [21]. Based on the uncontrolled discrete-event plant models and the requirements, the supervisory controller is synthesized using algorithms from the area of supervisory control theory [8, 27].
4. Simulation and simulation-based visualization can be used to validate the supervisory controller with respect to the uncontrolled hybrid plant.
5. The supervisory controller together with the uncontrolled discrete-event plant, called controlled system, can be subjected to verification (if there are desired properties that could not be expressed in terms of automata).
6. From the supervisory controller and the hybrid observer, an observer-based supervisor can be derived that takes care of translating command events to appropriate input signals for the actuators or set-points for feedback controllers, if needed. Based on this model, a real-time implementation can be generated.
7. Additionally, the supervisory controller model can be used for model-based testing [32], i.e., to generate test cases to which the real-time implementation can be subjected.

Important advantages of the proposed workflow are that a shift is made from developing supervisory controllers by hand to declarative modelling of the requirements that the supervisory controller should satisfy. As a consequence, adapting the supervisory controller due to changing requirements, which happens in any realistic industrial case, is relatively easy. Accordingly, the effort is put in modelling the system components and requirements and the corresponding supervisor expressed by automata is generated. Models of the supervisory controller obtained by synthesis can be analyzed by means of simulation and verification before being implemented, which increases confidence in correctness of the developed controllers. This way of working can be used for generation of supervisors for equipment, transport means (like Automated Guided Vehicles (AGV's)) but also for coordination control in a

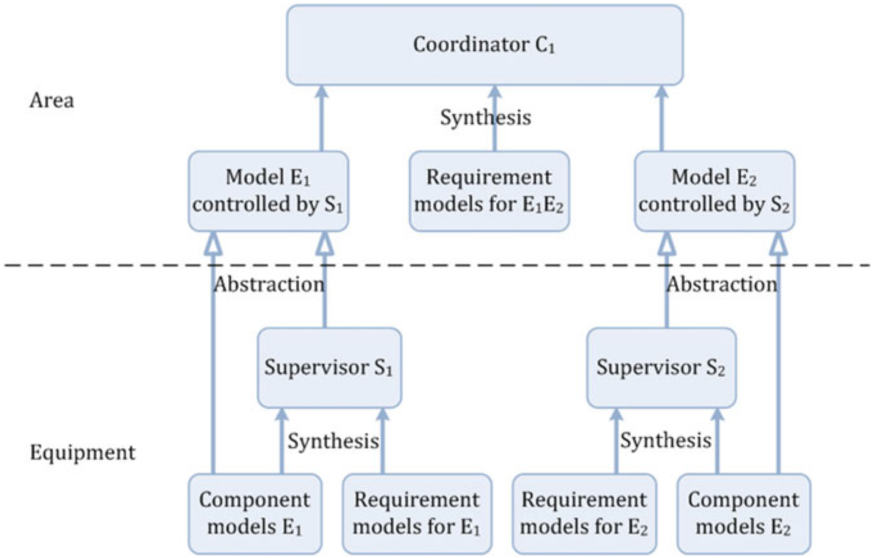


Fig. 3 A scheme for supervisory control synthesis

manufacturing area or even across different areas in a production facility, as schematically depicted in Fig. 3.

In the subsequent sections, we illustrate the first part of the workflow introduced above, that is steps 1 through 4. To this end, the multi mover case study introduced in [11] is used. As explained in [5, 25], autonomously navigating and cooperating vehicles (AGVs), of which the multi mover is an instance, are essential for a class of Cyber-Physical Production Systems.

3 Description of an AGV System

Automated Guided Vehicles (AGV) are driverless, battery-operated and computer-controlled vehicles that can transport materials within a manufacturing or distribution facility, as schematically shown in Fig. 4. These vehicles can be equipped with different types of steering and guidance. In this case study, an AGV called multi mover is considered which is equipped with one drive and one steer motor and which is able to follow a track integrated in the floor.

Every multi mover is equipped with several components that together need to take care of the given transport tasks. Additionally, the multi movers interact with each other and the supervisory controller that needs to be developed has to make sure that they all safely move around. In Sect. 3.1, the functionality of individual components is explained. Subsequently, in Sect. 3.2 the required interaction between the components is described.

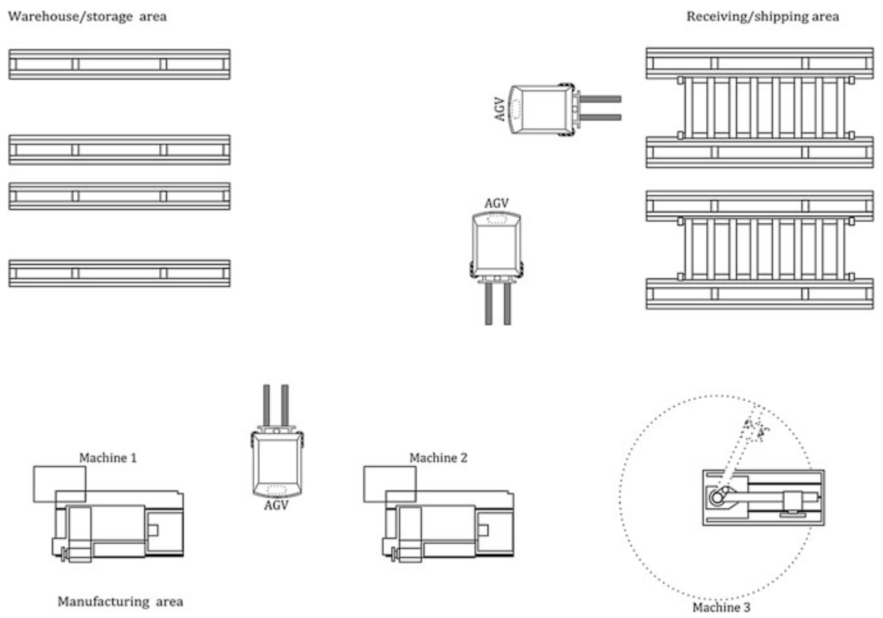


Fig. 4 Manufacturing system with AGV's

3.1 Components of the Multi Mover

A multi mover contains two different motors: a *drive motor* and a *steer motor*. The drive motor drives the multi mover forwards or backwards with a certain speed. The steer motor enables the multi mover to take turns so that it can follow the track integrated in the floor. The floor codes positioned near the track give additional information about it, such as the presence of a switch, a junction or a dead-end. To handle this information, the *ride control* component is integrated in the multi mover. The ride control receives the start and stop signal to make the multi mover start or stop riding, respectively.

As mentioned above, the multi mover is a battery-operated vehicle. The component that gives a signal if the battery level is too low is called the *battery sensor*. This is necessary to stop the multi mover safely instead of in an uncontrolled way when the battery is really empty.

To prevent collisions, the multi mover obtains information about obstacles, like walls, machines or other multi movers, through four *proximity sensors*. It is equipped with two pairs of sensors at the front and the back. Each of these pairs contains a sensor for long-range detection of 6 m and a sensor for short-range detection of 1 m.

There is a possibility that a moving object approaching the multi mover comes in direct contact with it. Therefore, the multi mover is equipped with a *bumper switch*, which signals a direct contact with an object. In such a case, the multi mover must stop directly because an unsafe situation could arise.

The multi mover is also equipped with three *LEDs* to display status information for the operator. The reset LED serves as an indication of errors. The other two LEDs, the forward and the backward LED, are used to indicate that the multi mover can be initiated in the corresponding direction.

The multi mover also has three *buttons* to operate the multi mover. One button is used to actuate the multi mover forwards, one to actuate it backwards and one to reset the multi mover when an error has occurred.

3.2 Interaction of the Components

The components described previously interact with each other. In Fig. 5, the interaction between the components is displayed graphically.

The track (wire) and the objects in the "Surroundings" are not part of the multi mover. However, they do interact with its components. The wire sends information to the multi mover like the start and stop signal of the ride control. The surroundings interact with the proximity sensors and the bumper switch because these sensors can be triggered by objects in the vicinity of the multi mover. The motors interact with the surroundings due to the fact that the multi mover moves.

The ride control interacts with the motors to control the movements. The proximity sensors interact with the motors, especially with the drive motor. If the long-range proximity sensor in the drive direction detects an object in the vicinity of 6 m, the drive motor slows down. If the short-range proximity sensor in the drive direction detects an object in the vicinity of 1 m, the drive motor stops. The battery sensor and the bumper switch also interact with the motors. The battery sensor and bumper switch shut down the motors if they are activated.

The buttons interact with the motors. The error button resets the multi mover to indicate that the errors are solved. The backward and forward button can initiate the multi mover in the corresponding direction. This has an influence on the drive motor. The LEDs are influenced by the motors and the battery sensor. When an

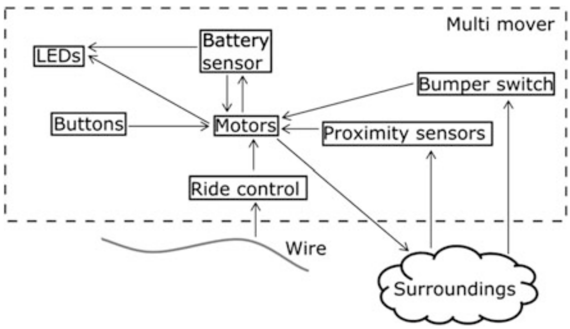


Fig. 5 Component interaction [28]. The *dashed box* indicates the system boundaries of the multi mover

error occurs in these components or when the battery sensor is activated, the reset LED is activated.

The to be developed supervisory controller has to assure that these components properly interact with each other to safely perform the required transport tasks.

4 Hybrid Models of the Uncontrolled System

In this section, the hybrid automata that represent the uncontrolled behavior of the system components are explained. To make these models not too complex for illustration purposes, the following two assumptions are made:

- The multi movers can only move in four directions, seen from above: left, right, up and down.
- Accelerations, in the driving speed as well as in the rotation speed, are infinite. In other words, the speed can change instantaneously.

Although physically unrealistic, it still gives a good understanding of the behavior of the system. Moreover, if needed more realistic models can be defined in the same set-up.

A hybrid automaton consists of locations and edges between those locations. Locations in combination with variables represent the states of the described components. Edges represent instantaneous transitions from one location to another (possibly the same) location. Guards may be used to model conditions under which such transitions may be enabled. Updates (also called assignments) are used to model value changes of variables as a consequence of taking a transition. Time passage and the associated continuous-time behavior of variables representing physical quantities (such as speed, position, force, etc.) is described by differential equations in the locations of the hybrid automata. Figure 6 gives an example of such a hybrid automaton.

Hybrid automata used in this chapter interact with each other by means of shared event synchronization and shared variables. In a network of automata, an event is only executed if all hybrid automata for which that event is defined are able to do so; otherwise it is disabled. Each variable is owned (and declared) by a single automaton and all other automata may read the value of this variable for use in guards and (right-hand sides of) updates of transitions. A typical property of the hybrid automata used is that event execution takes priority over time passage in a location.

The hybrid automata for the previously mentioned components are given in Fig. 6. They are explained in more detail in the following subsections.

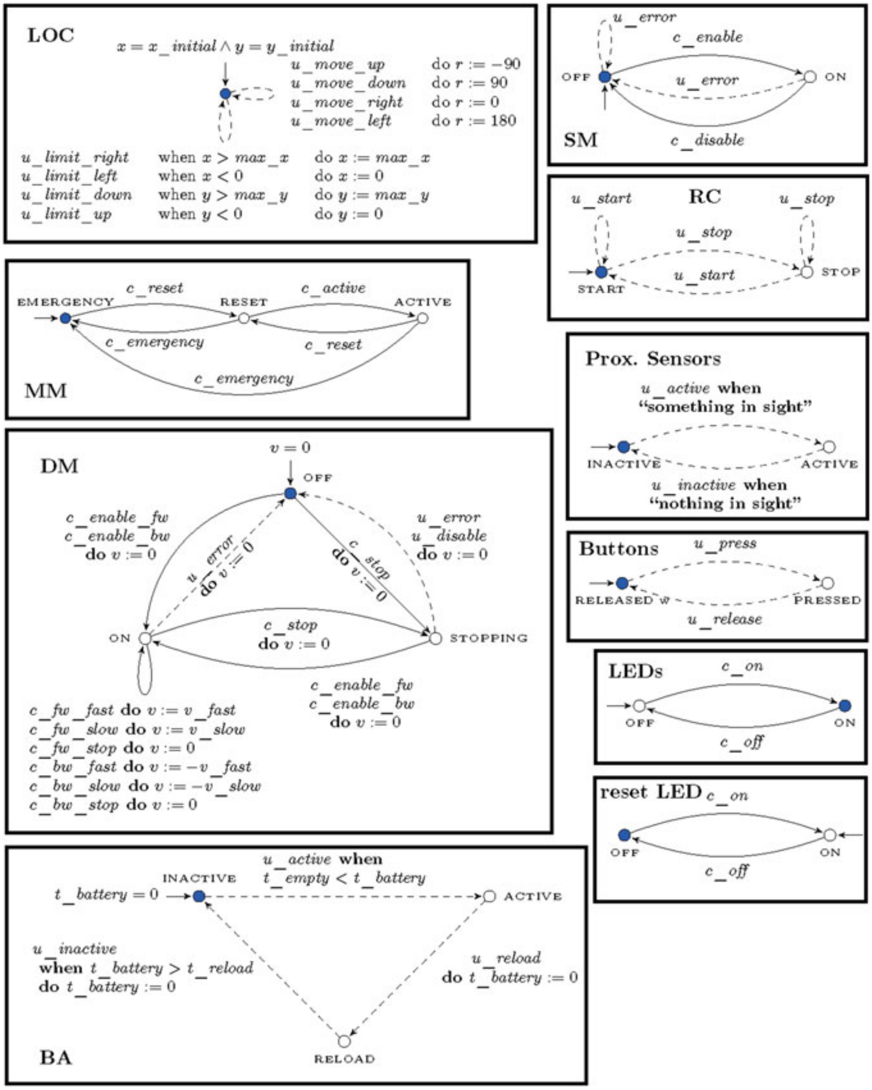


Fig. 6 Hybrid automata of the components of the multi mover. Differential equations that are used to define the values of the continuous-time variables are presented in the appropriate subsections

4.1 Multi Mover High-Level Modes and Movement

In Fig. 6, a model called LOC is introduced in which the physical location of the multi mover is defined. The hybrid automaton contains one location and several variables. Variables x and y with the associated derivatives are related to the horizontal and vertical coordinates of the vehicle. Variable r defines the rotational orientation of the multi mover in degrees. The front side of the multi mover is pointing in the direction of *r*. For the visualization the *x*, *y* and *r* are used to give the multi mover the right position and the right angle.

On the edge on the right side, the value of the variable *r* is defined depending on the steer command that is received from the environment. The derivatives of the location variables *x* and y depend on the rotation of the multi mover and the speed provided by the drive motor (*DM.v*). The physical behavior is described by the following differential equations:

$$\dot{x} = \text{if } r = 180\text{: } -DM.v \text{ elif } r = 0\text{: } DM.v \text{ else } 0;$$
$$\dot{y} = \text{if } r = -90\text{: } -DM.v \text{ elif } r = 90\text{: } DM.v \text{ else } 0;$$

For the visualization, it is assumed that the multi movers are moving in a rectangular area of dimensions given by the constants *max_x* and *max_y*. The boundaries of the area are formed by walls that can be detected by the proximity sensors.

The events on the edge below the state represent arriving at the boundaries of the ride area. The associated updates assure that the multi mover stays in this area.

The hybrid automaton of the multi mover (MM, see Fig. 6) is used to establish the high-level modes of the multi mover: *EMERGENCY*, *RESET* and *ACTIVE*. These are very useful when modelling the requirements in the next section.

The initial location is *EMERGENCY*, where all the systems of the multi mover must be disabled. It is the initial state because of the fact that the multi mover has to be initiated by the operator before it starts moving. The event *c_reset* should happen if the reset button is pushed.

In location *RESET*, the multi mover is reset and the multi mover can switch now to the *ACTIVE* state by the event *c_active*. There is also a possibility that there occurs an error, which triggers *c_emergency*.

In location *ACTIVE*, the multi movers systems are all enabled. When an error occurs the multi mover will go to the *EMERGENCY* location. The multi mover can also be reset with the *c_reset* event.

By means of the requirements defined in the next section, the events used in this high-level modes automaton are connected to events in the components.

4.2 Drive Motor

The model of the drive motor DM contains 3 locations: *OFF*, *ON*, and *STOPPING*. In location *OFF* the drive motor is disabled and cannot drive. In location *ON* it can move in different directions with different speeds. In location *STOPPING* it is stopped and can be disabled or enabled again. The model contains a variable *v* which represents the speed of the drive motor and so the speed of the multi mover. In Fig. 6, a graphical representation of the hybrid automaton is displayed.

OFF is the initial location because the multi mover starts with disabled motors. Initially the speed of the drive motor is also zero.

In this location, there are three possible events. The events to enable the drive motor, in the forward (*fw*) as well as in the backward (*bw*) direction. These events trigger a transition to location *ON*. There is also the possibility to stop the drive motor and go to location *STOPPING*.

In location *ON*, the events in the loop are the drive commands. It is also possible to go to location *STOPPING* with event *c_stop*.

In location *STOPPING*, the disable event of the drive motor, *u_disable*, can also occur. It is also possible to enable the drive motor again and go to the *ON* state. The uncontrollable event *u_error* models the occurrence of an error in the drive motor, this results in the *OFF* state of the drive motor. This event can occur in locations *ON* and *STOPPING*.

4.3 Steer Motor

The steer motor SM is simplified to a motor with locations *ON* and *OFF*. When the multi mover is moving and the steer motor must steer it in a certain direction (left, right or straight ahead), the steer motor must be *ON* to do this. In Fig. 6, a graphical representation of the hybrid automaton is displayed. The initial location is *OFF* because initially the multi mover is off and so is the steer motor.

There are three possible events in this model: *c_enable* and *c_disable* are controllable events that trigger the transition from *OFF* to *ON* and vice versa. Error occurrence is modelled with the event *u_error* which can always occur and results in a transition to the *OFF* state.

4.4 Ride Control

The ride control model (RC) contains two locations, *START* and *STOP*. Location *START* is initial because as soon as the multi mover is initiated by the operator it is the intention to let the multi mover drive. The model, which is graphically displayed in Fig. 6, contains the events *u_start* and *u_stop*. These events are uncontrollable because they are sent from the outside of the multi mover. The ride control always must be able to receive these signals.

4.5 Battery Sensor

In Fig. 6, a graphical representation of the model for the battery sensor BA is given. *INACTIVE* is the initial location because the system starts always with a loaded battery pack, which means that the sensor is inactive. Location *ACTIVE* represents the situation that the battery is empty and *RELOAD* that the battery is reloading.

The model contains several variables:

- *t_battery* indicates how much battery energy is used. The derivative of this variable represents how much energy is used at the moment. The used power is always minimally 1 because the multi mover is always using power and also depends on the speed of the multi mover.
- *t_empty* indicates how much energy the battery has initially. This is not a constant value because a battery is not always fully loaded and does not always deliver the same amount of energy.

The constant *t_reload* indicates how long it takes to reload the battery. The physical behavior of the sensor and its parameters is given by the following equations:

$$\dot{t\_battery} = 1 + \frac{speed^2}{9}$$
$$t\_empty = 3 * \sin\frac{time}{5*2.14} - \frac{1}{40} time + 200000$$

The event *u_active* contains a guard for when the battery runs empty. The event *u_reload* is controlled by the operator which is responsible for reloading the multi movers. In reality this is a physical action. The event *u_inactive* represent the deactivation of the sensor when the battery is fully loaded, the guard assures this behavior.

4.6 Proximity Sensor

A graphical representation of the hybrid automata of the proximity sensors can be seen in Fig. 6. The proximity sensor has to two locations: *ACTIVE* and *INACTIVE* with transitions with the events *u_active* and *u_inactive* between them. The guard "something in sight" makes sure that the event only occurs when there a multi mover or a wall is in the vicinity of that specific proximity sensor. The guard is expressed in terms of the position of the multi mover and the positions of the walls and the other multi mover(s), and are left out for conciseness. The guard "nothing in sight" expresses that it is not the case that something is in sight.

4.7 Bumper Switch

A graphical representation of the hybrid automaton of the bumper switch BS is displayed in Fig. 6. Also the bumper switch can only be active or inactive and there are two uncontrollable events which also contain a guard to ensure the physical behavior. The guards "Bumper touched" and "Bumper untouched" characterize whether there is a multi mover or a wall that touches the bumper or not, and are left out.

In order to model unexpected objects that may activate the bumper switch an additional hybrid automaton is introduced that captures these uncontrollable events. It is a copy of the bumper switch automaton with the guards removed.

4.8 LEDs and Buttons

Each of the LED models contains two locations: *ON* and *OFF*. These locations are connected with the events *c_off* and *c_on*. A graphical representation of the hybrid automata is displayed in Fig. 6.

The backward and forward LEDs indicate whether the multi mover can be initiated or not. This is initially not the case. Therefore, the initial state is the *OFF* state.

The reset LED of the multi movers has a slightly different model. The locations and events are the same, only the initial location is *ON* because the system starts in the error state and must be restarted.

The hybrid automaton of a button also has two states and two transitions. The button is either pressed or released and the possible events are to press or release it. However, the events are uncontrollable because they are caused by the operator of the multi mover and therefore uncontrollable for the supervisor.

4.9 Abstraction Form Hybrid Automata to Discrete-Event Models

In order to synthesize a supervisor with the CIF tool set, in Sect. 6, a discrete event model of the uncontrolled system is needed. This means that the hybrid plant which has been introduced above needs to be abstracted into a discrete-event model. This is achieved by removing all continuous variables. To this end, their declarations are removed, the differential equations in which their evolution over time is defined are removed, all updates in which they appear on the left-hand side are removed, and all parts of guards in which they are evaluated are replaced by true. As an example, the discrete-event automaton obtained for the battery sensor BA from Fig. 6 is given in Fig. 7.

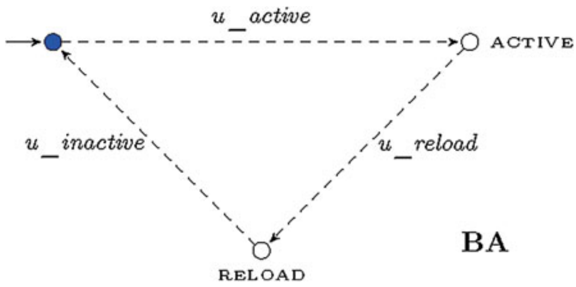


Fig. 7 Discrete-event automaton for the battery sensor

5 Requirements of the System

There are several requirements to make sure that every multi mover safely and correctly performs its transport tasks. These requirements can be used to synthesize a supervisor, which is explained in the subsequent section.

5.1 *Emergency and Error Handling*

In relation to the emergency and the error handling, the following requirements are defined. The multi mover (*MM*) is only allowed to go to the *RESET* and *ACTIVE* state when the battery is inactive and nothing touches the bumper.

$$\rightarrow \{MM.c\_reset, MM.c\_active\} \Rightarrow BA.INACTIVE \downarrow \wedge BS.INACTIVE \downarrow$$

The events *c_reset* and *c_active* correspond with going to the given states. The *BA.INACTIVE* and the *BS.INACTIVE* states correspond to the given states of the battery sensor and the bumper switch.

The multi mover needs to be reset after an error event has occurred. These error events are an error in the steer motor (*SM*) or drive motor (*DM*), a battery which runs empty, or a bumper switch which is pressed. An emergency may only be issued when an error situation has occurred. The automaton in Fig. 8 describes these

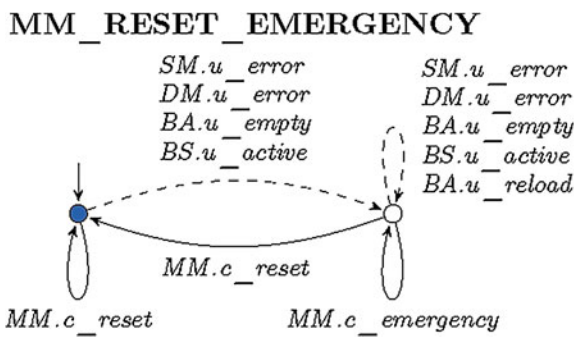


Fig. 8 Requirement automaton for emergency and error handling

requirements. Due to the fact that there is a certain sequence in this requirement, this cannot be stated as a state-based expression.

5.2 LED Actuation

The requirements related to LED actuation are the following.

- The reset LED (*RL*) may only be switched off if the multi mover is in the *ACTIVE* or *RESET* state, because in these states there is no error.

$$\rightarrow \{RL.c\_off\} \Rightarrow MM.ACTIVE\downarrow \vee MM.RESET\downarrow$$

- The reset LED may only be turned on when the multi mover is in the *EMERGENCY* state, because then an error has occurred.

$$\rightarrow \{RL.c\_on\} \Rightarrow MM.EMERGENCY\downarrow$$

- Forward LED (*FL*) and the backward LED (*BL*) may only be switched on if the multi mover is in its *RESET* state.

$$\rightarrow \{FL.c\_on,\ BL.c\_on\} \Rightarrow MM.RESET\downarrow$$

- The forward and backward LED may only be switched of if the multi mover is in the *ACTIVE* or *EMERGENCY* state. This due to the fact that in these cases the multi mover cannot be initiated and therefore the LED should be able to switch off.

$$\rightarrow \{FL.c\_off,\ BL.c\_off\} \Rightarrow MM.ACTIVE\downarrow \vee MM.EMERGENCY\downarrow$$

5.3 Motor Actuation

The drive motor may only be fully stopped if the multi mover is in its reset or emergency mode and if the scene program handler is off. The multi mover should be in one of these modes because the drive motor may be disabled after stopping the drive motor. The scene program handler should be off because then no drive commands can be received.

$$\rightarrow \{DM.c\_stop\} \Rightarrow (MM.RESET\downarrow \vee MM.EMERGENCY\downarrow) \wedge SH.OFF\downarrow$$

The drive motor may only be stopped if it was driving forward or backwards. For establishing the drive direction the observer for the drive direction is used

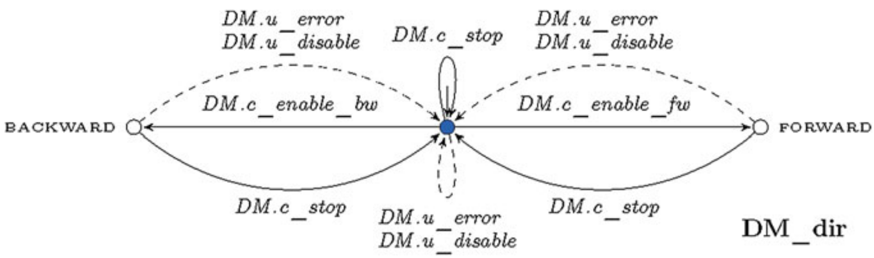


Fig. 9 Hybrid automaton for the observer of the drive motor direction

(*DM_dir*); for this model see Fig. 9. The events *DM.u_error*, *DM.u_disable* and *DM.c_stop* can always occur and will always result in the *STANDSTILL* state. In the *STANDSTILL* state the events *DM.c_enable_fw* and *DM.c_enable_bw* are resulting in the states *FORWARD* and *BACKWARD*, respectively.

$$\rightarrow \{DM.c\_stop\} \Rightarrow DM\_dir.FORWARD \downarrow \vee DM\_dir.BACKWARD \downarrow$$

The steer motor may only be disabled if the multi mover is in its reset or emergency mode and the drive motor is off. This is important because this guarantees that the multi mover is at a full stop and an error occurred before the steer motor is disabled.

$$\rightarrow \{SM.c\_disable\} \Rightarrow (MM.RESET \downarrow \vee MM.EMERGENCY \downarrow) \wedge DM.OFF \downarrow$$

The steer motor may only be enabled when the multi mover is active because only then the multi mover has the intention to drive and only then the steering is necessary.

$$\rightarrow \{SM.c\_enable\} \Rightarrow MM.ACTIVE \downarrow$$

The drive motor can only be enabled backward or forward when the multi mover is active and the steer motor is on:

$$\rightarrow \{DM.c\_enable\_fw, DM.c\_enable\_bw\} \Rightarrow MM.ACTIVE \downarrow \wedge SM.ON \downarrow$$

The drive motor should only be able to execute drive commands if the multi mover is *ACTIVE*:

$$\rightarrow \left\{ \begin{array}{c} DM.c\_fw\_slow, DM.c\_bw\_slow, DM.c\_fw\_fast, \\ DM.c\_bw\_fast \end{array} \right\} \Rightarrow MM.ACTIVE \downarrow$$

5.4 Button Handling

The following requirements are related to the button handling of the multi mover. The multi mover may only be activated if either the forward button (*FB*) or the backward button (*BB*) is pressed, and the reset button (RB) is not pressed:

$$\rightarrow \{MM.c\_active\} \Rightarrow \neg\, RB.PRESSED \downarrow \wedge$$
$$((FB.PRESSED \downarrow \wedge\ BB.RELEASED \downarrow) \vee (BB.PRESSED \wedge FB.RELEASED \downarrow))$$

The multi mover may only be reset if the reset button is pressed by the operator:

$$\rightarrow \{MM.c\_reset\} \Rightarrow RB.PRESSED \downarrow$$

The requirements in Fig. 10 state that when the forward or backward button is pressed, first the multi mover is enabled and only then the drive motor is enabled in the direction corresponding to the pressed button.

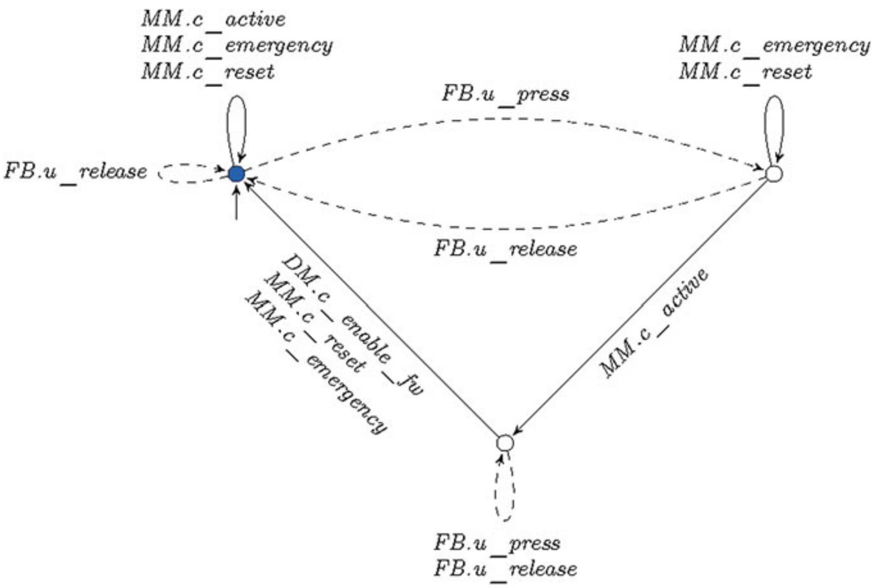


Fig. 10 Requirements for button handling. Shown is the automaton for handling of the forward button. A similar requirement automaton is needed for the handling of the backward button

5.5 Proximity Sensors and Ride Control Handling

The drive motor may only stop driving forward/backward as the ride control (*RC*) is in *STOP* state or the short-range proximity sensor on the front/back side is *ACTIVE*.

$$\rightarrow \{DM.c\_fw\_stop\} \Rightarrow RC.STOP \downarrow \vee\ FSP.ACTIVE \downarrow$$
$$\rightarrow \{DM.c\_bw\_stop\} \Rightarrow RC.STOP \downarrow \vee\ BSP.ACTIVE \downarrow$$

The drive motor may only be stopped when the multi mover is moving in the corresponding direction. The state-based expression uses the variables introduced by the observer *DM_speed* in Fig. 11.

$$\rightarrow \{DM.c\_fw\_stop\} \Rightarrow \begin{pmatrix} DM\_speed.fw\_fast \vee DM\_speed.fw\_slow \\ \vee\ DM\_speed.bw\_fast \vee DM\_speed.bw\_slow \end{pmatrix} \wedge DM\_dir.FORWARD \downarrow$$
$$\rightarrow \{DM.c\_bw\_stop\} \Rightarrow \begin{pmatrix} DM\_speed.fw\_fast \vee DM\_speed.fw\_slow \\ \vee\ DM\_speed.bw\_fast \vee DM\_speed.bw\_slow \end{pmatrix} \wedge DM\_dir.BACKWARD \downarrow$$

The observer of the drive commands is a model with one single state. The model contains four Booleans, for every drive command one. As the event occurs the corresponding Boolean is set to the value true and the others to false. This way the last given drive command will be known. The events below the state set all the Booleans to false because the drive motor is not driving after such a command. All the transitions in this model are linked to the drive motor model.

The following requirements state basic conditions for some of the events of the drive motor:

- the drive motor may only go forward/backward if the ride control is in *START* state and the front/back side short-range proximity sensor is *INACTIVE*

$$\rightarrow \{DMc\_fw\_fast, DM.c\_bw\_slow\} \Rightarrow RC.START \downarrow \wedge\ FSP.INACTIVE \downarrow$$
$$\rightarrow \{DMc\_bw\_fast, DM.c\_bw\_slow\} \Rightarrow RC.START \downarrow \wedge\ BSP.INACTIVE \downarrow$$

- the multi mover may only start driving slow as the long-range proximity sensor is ACTIVE and fast when the long-range proximity sensor is INACTIVE

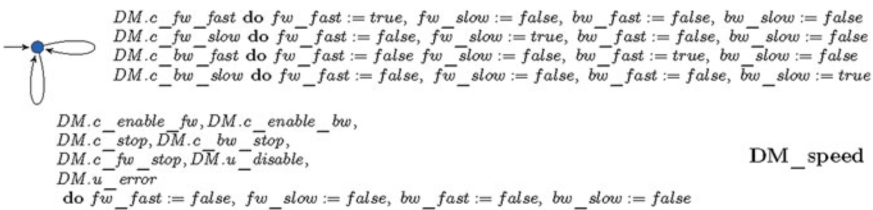


Fig. 11 Automaton for observer of drive commands

$$\rightarrow \{DM.c\_fw\_slow\} \Rightarrow FLP.ACTIVE \downarrow$$
$$\rightarrow \{DM.c\_fw\_fast\} \Rightarrow FLP.INACTIVE \downarrow$$
$$\rightarrow \{DM.c\_bw\_slow\} \Rightarrow BLP.ACTIVE \downarrow$$
$$\rightarrow \{DM.c\_bw\_fast\} \Rightarrow BLP.INACTIVE \downarrow$$

- forward/backward driving may only occur when the multi mover is moving in the right direction and a single drive command may not occur several times in a row

$$\rightarrow \{DM.c\_fw\_fast\} \Rightarrow DM\_dir.FORWARD \downarrow \wedge \neg DM\_speed.fw\_fast$$
$$\rightarrow \{DM.c\_fw\_slow\} \Rightarrow DM\_dir.FORWARD \downarrow \wedge \neg DM\_speed.fw\_slow$$
$$\rightarrow \{DM.c\_bw\_fast\} \Rightarrow DM\_dir.BACKWARD \downarrow \wedge \neg DM\_speed.bw\_fast$$
$$\rightarrow \{DM.c\_fw\_slow\} \Rightarrow DM\_dir.BACKWARD \downarrow \wedge \neg DM\_speed.bw\_slow$$

6 Synthesis of Supervisory Controller

In this section, a supervisor is synthesized for a system consisting of two multi movers.

Supervisory control theory [27] defines the construction of a model of a supervisory controller based on a discrete-event model of the plant and the requirements. The resulting supervisor not only satisfies the requirements but also the following properties:

- The controlled system (i.e., plant and supervisor together) is *non-blocking*. A system is considered nonblocking when from each reachable state a so-called marked state may be reached.
- *Controllability*, i.e., the supervisor has not disabled any uncontrollable events.
- *Maximal permissiveness* w.r.t. the plant, i.e., the supervisor only disables events that result in violation of the requirements or the previous properties.

Originally, supervisory control theory was only capable of dealing with finite automata for both plant and requirements. Later, the more sophisticated (though equal in expressivity) state-based requirements have been introduced [20, 21], and extensions of the theory to deal with extended finite automata (in which variables are used) have been proposed [23] and implemented in tools such as Supremica [2] (http://www.supremica.org) and CIF (http://www.cif.se.wtb.tue.nl).

In the models of the discrete-event and hybrid automata in Sect. 4 and in the requirement automata in Sect. 5, the marked locations are indicated by a filled node. Unless stated otherwise, for the variables all values are considered to be marked. This is the case in this case study.

A supervisor has been synthesized for a system consisting of two multi movers. Therefore, all models and requirements should be applied to both multi movers. For this the, so-called, data-based synthesis option of the CIF toolset has been used. Synthesis of the supervisor took approximately half a second with a 64 bit computer with an Intel Core i5vPro processor.

To indicate the size and complexity of the system and the input of the synthesis, an overview of some of the size characteristics of the system are provided. The uncontrolled plant consists of 34 automata (17 per multi mover). There are 10 automata-based requirements (5 per multi mover), and 58 state-based requirements (29 per multi mover). Each automaton has from 1 to 3 locations, maximally 9 controllable and 8 uncontrollable events, and between 2 and 15 edges.

The resulting supervisor consists of the 34 plant automata, the 10 automata-based requirements, a supervisor automaton for each of the 58 state-based requirements and one additional supervisor automaton with one state and an edge for each of the 40 controllable events of the system representing the additional constraints needed to obtain a proper supervisor. In this case study, it turns out that the additional constraints are trivial (equivalent to *true*) in each case.

Alternatively to the approach presented before, one can synthesize a local supervisor for each of the multi movers. The multi movers cannot communicate with each other. Therefore, it is a logical decision to provide each multi mover with a separate supervisor. These separate supervisors are faster to synthesize (since the problem is approximately half of the size) and in general easier to understand. A potential problem with this approach is that the obtained supervisors may be conflicting, i.e., when combined result in a controlled system in which blocking occurs. In this particular case, the combination of the local supervisors of the two multi movers results in an identical supervisor as was obtained previously. Using the available tooling of the CIF toolset it is possible to compute the complete state space of this supervisor, which in this case has more than 1.7 million states and over 43 million transitions.

7 Simulation-Based Visualization

Although the supervisor has been synthesized based on a discrete-event model of the uncontrolled system and formalized requirements, it is still needed to validate the proper functioning of this supervisor in its hybrid context. It could, for example, be the case that the abstraction from hybrid models to discrete-event models has resulted in a supervisor that does not function as expected. Or, one could have made a mistake in some of the requirements. In the proposed model-based engineering framework, this is achieved by means of simulation-based visualization.

For the purpose of simulation-based visualization, an interactive graphical set-up is created using SVG. This is a graphical representation in which the modelled system is visualized and in which state information of the controlled system may be shown to the user. For the visualization a rectangular surface is chosen with a length of 60 m and a width of 40 m. For simplicity of the visualization, only two multi movers, which drive independently, are considered. If there would be more multi movers, the visualization would be unclear, large and chaotic because for every multi mover there should be space for buttons and information. Even with this assumption, relevant and representative situations can still be simulated. It is

Fig. 12 Graphical set-up for the simulation-based visualization

assumed that only multi movers and walls can be detected by the proximity sensors and bumper switch. For unexpected collisions with the multi mover a button is implemented in the interactive visualization. The interactive visualization graphical figure is displayed in Fig. 12.

The figure contains several parts. In the middle, the drive area for the multi movers is displayed in gray with two multi movers on it. The area is surrounded by walls. The red circles represent the multi movers. The multi movers are red when they are inactive or in the error or emergency state and green if they are active. The small arrows on the multi movers indicate the front side of the multi movers. The letters indicate which multi mover it is, A or B.

The left-hand side of the visualization figure represents the actual interfaces of the multi movers. The LEDs and the buttons correspond to the LEDs and the buttons in the models. During simulation, these buttons are used to initiate the multi movers.

Left from the drive area, the states of the sensors and motors are displayed. For example, the current speed of the multi mover is displayed with behind it, the state of the drive motor. Based on such state information, it can be seen if the sensors react as expected and what happens in the system.

Right from the drive area, the following groups can be found from top to bottom:

1. The interactive buttons that represent sending signals to the ride control. These buttons can be used to send a start or stop signal to the multi movers. The letter indicates the multi mover the buttons correspond to.
2. The interactive buttons related to movement definition. The four buttons around A and around B can be used to send desired steer commands to the multi movers. The big buttons with the two arrows can be used to change the drive direction of the multi movers from forward to backward and vice versa.
3. The interactive buttons representing the physical handling of the operator reloading the batteries of the multi movers. When a battery is empty the letter turns red. By clicking the button the battery of the multi mover will be reloaded. During reloading the button is green. As soon as the loading process is

completed the button will turn gray. When the battery has enough energy to let the multi mover drive, the letter is green.

4. The interactive buttons for simulating errors in the system. Clicking one of them introduces an error. Also the bumper switch can be activated. Normally these events happen unexpectedly. However, to see what happens in the simulation the errors can be introduced to check if the system responds to these unexpected events as it should.

8 Concluding Remarks

In this chapter, we have illustrated a model-based engineering process for the development of a supervisory controller. Parts of this approach have been applied to industrial cases in the past few years involving amongst others lithography machines [36], baggage handling systems [17], MRI scanners [31], automotive systems [35], electron microscopes [16] and container terminal systems [37].

The demonstrated approach of model-based systems engineering in combination with supervisory control and an interactive visualization is experienced as a good way to represent a system and to retrieve a good idea of the behavior of a system relatively fast. Due to the visualization it is easy to create a reasonable experience of how the system works and how it reacts to certain situations. This way also people that are relatively unknown to the system can easily understand how the system works.

Although not illustrated in this chapter, the proposed way of working also allows to take into consideration verification, code generation, and model-based testing to some extent. To facilitate verification the CIF toolset has model transformations (of proper subsets) to the model checkers mCRL2 [9] and UPPAAL [19]. Code generation to PLC code has been implemented and used in the context of the case studies involving baggage handling systems [30]. Model-based testing is currently not supported by the CIF toolset, but is investigated in several projects [1].

References

1. Aerts A, Mousavi M, Reniers M (2015) A tool prototype for model-based testing of cyber-physical systems. In: ICTAC 2015. Springer, pp 563–572
2. Akesson K, Fabian M, Flordal H, Malik R (2006) Supremica—an integrated environment for verification, synthesis and simulation of discrete event systems. In: WODES
3. Baeten JCM, Basten T, Reniers MA (2009) Process algebra: equational theories of communicating processes. Cambridge University Press, Cambridge
4. Becker S, Dziwok S, Gerking C, Heinzemann C, Thiele S (2014) The MechatronicUML design method-process and language for platform-independent modeling. Software Engineering Group, Heinz Nixdorf Institute, University of Paderborn

5. Böckenkamp A, Weichert F, Stenzel J, Lünsch D (2016) Towards autonomously navigating and cooperating vehicles in cyber-physical production systems. In: Machine learning for cyber-physical systems, pp 112–120
6. Brettel M, Friederichsen N, Keller M, Rosenberg N (2014) How virtualization, decentralization and network building change the manufacturing landscape: an industry 4.0 perspective. Int J Sci Eng Technol 8(1):37–44
7. Bringmann E, Krämer A (2008) Model-based testing of automotive systems. In: International conference on software testing, verification, and validation, pp 485–493
8. Cassandras CG, Lafortune S (2008) Introduction to discrete event systems. Springer, New York
9. Cranen S, Groote JF, Keiren JJ, Stappers FP, de Vink EP, Wesslink JW, Willemse TAC (2013) An overview of the mCRL2 toolset and its recent advances. In: TACAS, pp 199–213
10. Estefan JA (2008) Survey of model-based systems engineering (MSBE) methodologies. In: INCOSE
11. Forschelen ST, van de Mortel-Fronczak JM, Su R, Rooda JE (2012) Application of supervisory control theory to theme park vehicles. Discrete Event Dyn Syst 22(4):511–540
12. Friedenthal S, Moore A, Steiner R (2006) OMG systems modeling language (OMG SysML™). In: INCOSE international symposium
13. Fritzson P, Bunus P (2002) Modelica—a general object-oriented language for continuous and discrete-event system modeling and simulation. In: 35th IEEE annual simulation symposium
14. Henzinger TA (1996) The theory of hybrid automata. In: LICS, pp 278–292
15. Isermann R, Schaffnit J, Sinsel S (1999) Hardware-in-the-loop simulation for the design and testing of engine-control systems. Control Eng Pract 7(5):643–653
16. Janssen T, van de Mortel-Fronczak JM, van Gerwen E, Reniers MA (2015) Synthesis of discrete-event controllers from sequence-based. In: CDC
17. Kamphuis RH (2013) Design and real-time implementation of a supervisory controller for baggage handling at Veghel airport. Master thesis, Systems Engineering, Mechanical Engineering, TU/e
18. Khaitan S, McCalley J (2015) Design techniques and applications of cyberphysical systems: a survey. IEEE Syst J 9(2):350–365
19. Larsen KG, Pettersson P, Yi W (1997) UPPAAL in a nutshell. Int J Softw Tools Technol Transf 1(1–2):134–152
20. Ma C, Wonham W (2005) Nonblocking supervisory control of state tree structures. Springer, Berlin
21. Markovski J, Jacobs KG, van Beek DA, Somers LJ, Rooda JE (2010) Coordination of resources using generalized state-based requirements. In: WODES, pp 287–292
22. National Science Foundation (2015, November 15) Cybermanufacturing systems. Retrieved from http://www.nsf.gov/pubs/2015/nsf15061/nsf15061.jsp
23. Ouedraogo L, Kumar R, Malik R (2011) Nonblocking and safe control of discrete-event systems modeled as extended finite automata. IEEE Trans Autom Sci Eng 8(3):550–569
24. Peterson JL (1977) Petri nets. ACM Comput Surv 9(3):223–252
25. Peterson T, Schindel W (2015) Model-based system patterns for automated ground vehicle platforms. In: INCOSE international symposium 2015, pp 388–403
26. Pulecchi T, Casella F (2008) HyAuLib: modelling hybrid automata in modelica. In: Modelica
27. Ramadge PJ, Wonham WM (1987) Supervisory control of a class of discrete event processes. SIAM J Control Optim 25(1):206–230
28. Roelofs KW (2015) Simulation-based visualization for multi movers. Eindhoven: Bachelor thesis, Systems Engineering, Mechanical Engineering, TU/e
29. Skoldstam M, Akesson K, Fabian M (2007) Modeling of discrete event systems using finite automata with variables. In: CDC, pp 3387–3392
30. Swartjes L, van Beek DA, Reniers MA (2014) Towards the removal of synchronous behavior of events in automata. In: WODES, pp 188–194

31. Theunissen RJ, Petreczky M, Schiffelers RR, van Beek DA, Rooda JE (2014) Application of supervisory control synthesis to a patient support table of a magnetic resonance imaging scanner. IEEE Trans Autom Sci Eng 11(1):20–32
32. Utting M, Legeard B (2007) Practical model-based testing: a tools approach. Morgan Kaufmann, San Francisco
33. van Beek DA, Fokkink WJ, Hendriks D, Hofkamp A, Markovski J, van de Mortel-Fronczak JM, Reniers MA (2014) CIF 3: model-based engineering of supervisory controllers. In: TACAS, pp 575–580
34. van de Mortel-Fronczak JM, Gorter RJ, Rooda JE (2000) A case study in simulation-based system specification and design. In: ESM, pp 232–238
35. van de Mortel-Fronczak JM, Huisman RG, van der Heijden MH, Reniers MA (2014) Supervisor synthesis in model-based automotive systems engineering. In: ICCPS, pp 187–198
36. van der Sanden B, Reniers MA, Geilen M, Basten T, Jacobs J, Voeten J, Schiffelers RR (2015) Modular model-based supervisory controller design for wafer logistics in lithography machines. In: MODELS, pp 416–425
37. Volmer B (2015) Supervisory controller design for a container terminal system. Master thesis, Control Systems Technology, Mechanical Engineering, TU/e

Formal Verification of SystemC-based Cyber Components

Daniel Große, Hoang M. Le and Rolf Drechsler

1 Introduction

Cyber-Physical Systems (CPS) integrate physical and cyber components. These cyber components (e.g. HW/SW implementation of embedded control algorithms) are responsible for the computation part of CPS. Developing such complex components within todays time-to-market constraints requires building abstract models for architectural exploration and early software development. This procedure has been systematized resulting in the *Electronic System Level* (ESL) design [2]. For ESL design, *SystemC* [3] has become the standard modeling language and is nowadays being employed in various industries (including consumer electronics, automotive, industrial automation, etc.). SystemC is a C++ class library and provides modules, ports, interfaces and channels as the fundamental modeling components, whereas, the functionality is described by processes. In addition, the SystemC library also includes an event-driven simulation kernel. Essentially, the simulation kernel executes the processes non-preemptive and manages their parallel execution by using delta-cycles.

Most crucial for the success of SystemC is the concept of *Transaction Level Modeling* (TLM) [4–6]. TLM enables the description of communication in terms of

---

This chapter is an extended version of a conference paper appeared at MEMOCODE 2010 [1].

D. Große (✉) · H.M. Le · R. Drechsler
University of Bremen, Bremen, Germany
e-mail: grosse@informatik.uni-bremen.de

H.M. Le
e-mail: hle@informatik.uni-bremen.de

R. Drechsler
e-mail: drechsle@informatik.uni-bremen.de

D. Große · R. Drechsler
DFKI Bremen, Bremen, Germany

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_6

abstract operations (transactions). The simulation of SystemC TLM models is orders of magnitude faster in comparison to synthesizable HW models, which are implemented at *Register Transfer Level* (RTL) using e.g. VHDL or Verilog. Furthermore, TLM allows interoperability between models from different IP vendors. Please note that this chapter does not target any particular TLM library (e.g. TLM-2.0 standard) but handles TLM in a more general sense: we focus on SystemC TLM designs where communication is done through transactions (i.e. calling interface functions) and the synchronization is based on events.

Clearly, an abstract SystemC TLM model provides the first formalization of the design specification. This first TLM model is usually untimed and is successively refined by adding timing information to a timed TLM model, which in turn is refined down to RTL. Therefore, potential bugs need to be identified already at TLM. However, this functional verification task is difficult [7]. Methods commonly applied at TLM rely on simulation (see e.g. [8–13]) and therefore cannot guarantee the functional correctness, which is of uttermost importance, if the cyber components under development will be integrated in safety-critical CPS.

The existing formal verification approaches for SystemC TLM designs mainly check properties local to processes or have extremely high run-time (more details are discussed in the related work section). Hence, they cannot be used to verify major TLM behavior such as the start of a transaction after a certain event. In contrast, the approach proposed in this chapter makes the following contributions:

- *Verification of "true" TLM properties*: In addition to simple safety properties the user can check the effect of transactions and the causal dependency between events and transactions. The *Property Specification Language* (PSL) is used to formulate a property.
- *Adjustment of temporal resolution*: The approach allows to specify the sampling rate of the temporal operators, e.g. the user can focus on certain events or start/end of specific transactions.
- *Automated verification method*: The approach performs a fully automatic SystemC-to-C transformation. Then, monitoring logic for the property is automatically embedded into the C model. This monitoring logic uses C assertions and *Finite State Machines* (FSMs). To verify the property, the verification method of *Bounded Model Checking* (BMC) is employed on the C model.
- *Efficiency and completeness*: An induction-based verification method working on the level of C is proposed for the generated models, making the approach complete and much more efficient.

For different SystemC TLM designs we report the verification of properties describing important behavior at TLM which has not been possible before. Moreover, the experiments demonstrate that complete proofs can be carried out efficiently using induction.

This chapter extends our previous work [1] by supporting timed SystemC constructs and providing a much more precise and formal description of the SystemC-to-C transformation. Additionally, much more experimental results, especially a comparison with the up-to-date related approach ESST [14], have been included. The proposed techniques have been implemented as the tool *SCIVER* which stands for *S*ystem*C I*nduction-based *VER*ifier for Transaction Level Models. The benchmarks (i.e. SystemC TLM model, properties, generated C model) can be downloaded from our website.[1]

This chapter is structured as follows: In Sect. 2 related work is discussed. The preliminaries are provided in Sect. 3. Section 4 introduces the TLM property checking approach. First, a simplified SystemC kernel is presented. Based on this kernel the automatic generation of the verification model is introduced. Then, the property language and the generation of the respective monitors are given. In the last part of this section the BMC-based verification technique is presented. Section 5 describes the induction-based verification method for the transformed models. The experimental evaluation is presented in Sect. 6. Finally, the chapter is summarized and ideas for future work are outlined.

2 Related Work

One of the first formal approaches for SystemC TLM verification has been introduced in [15]. However, the design entry of this method is UML and only during the construction of the derived FSM some properties can be checked.

Another approach has been proposed in [16]. A formal model can be extracted in terms of communicating state machines and can be translated into an input language for several verification tools. Simple properties on very small designs have been verified with this approach.

The authors of [17] translate a SystemC design into Petri nets and then apply CTL model checking. However, the resulting Petri nets become very large even for small SystemC descriptions as the experiments have shown.

In [18] a technique has been presented that allows CTL model checking for SystemC TLM designs, but the SystemC TLM design has to be transformed manually to a dedicated automata model.

The approach of [19] translates a SystemC TLM design into Promela. The Promela model is then checked by the model checker SPIN. The translation is entirely manual and properties related to events and transactions are not considered.

A translation of untimed SystemC TLM to the process algebra LOTOS is proposed in [20]. However, the focus of the work is to compare two possible LOTOS encodings, property checking is not discussed.

In [21] an approach mapping SystemC designs into UPPAAL timed automata has been proposed. It differs from our approach in particular regarding the expressiveness of the properties—the properties have to be specified on the UPPAAL model and sometimes helper automata need to be defined.

[1]www.systemc-verification.org/sciver.

The work in [22] presented an approach combining static and dynamic *Partial Order Reduction* (POR) techniques to detect deadlocks and simple safety property violations. Blanc and Kroening [23] proposed a static POR technique for state space exploration of SystemC designs using model checking, but property checking was not considered. The limitations of both approaches are that representative inputs need to be provided and the absence of corner-case errors cannot be proven.

Most closely related to our proposed approach is the software model checking approach for SystemC introduced in [24]. Both approaches translate SystemC designs to sequential C programs using very similar mechanisms. Then, available C model checkers can be applied directly to the resulting C programs. Both approaches also include novel verification techniques which outperform this straight-forward solution significantly. Nevertheless our approach differs from [24] clearly in two key aspects.

- First, the employed verification techniques are different. In [24], explicit-state model checking has been combined with abstraction-based symbolic techniques to deal with the SystemC scheduler and the execution of SystemC processes, respectively. Further enhancement incorporating POR techniques has been also proposed in [14]. In contrast, we apply a novel high-level induction proof over the scheduler loop to the generated C models. In our experiments we also provide a comparison of our approach and [14].
- Second, there is a clear distinction in the expressiveness of the properties. The approach of [24] (and [14]) only considers local C assertions, whereas we support high-level temporal properties with TLM primitives. Note that while these high-level properties do not bring additional challenges to backend model checkers, it is mandatory for any practical model checking solution aiming SystemC TLM designs to support them—manually writing local assertions to check high-level behaviors is a very tedious and error-prone process. Moreover, it is also possible to use their approach as a verification back-end of our approach.

With regard to property languages for SystemC, a fundamental work has been published [25]. The authors define a trace semantic for SystemC covering also abstract models. Furthermore, a PSL [26] oriented language has been introduced which additionally includes new primitives to allow expressing software aspects like for example pre- or post-conditions. We use the introduced PSL primitives in this work. The respective details are discussed in Sect. 4.

Recently complementary formal verification techniques for SystemC based on symbolic simulation have been proposed in [27–29]. They combine symbolic execution with complete exploration of all process schedules. An in-depth comparison as well as a combination of the best of both worlds is left for future work.

3 Preliminaries

This section provides essential background to help understand the formal verification approach proposed in this chapter. First, a brief overview of Bounded Model Checking and Induction is given in Sect. 3.1. Then Sect. 3.2 introduces the basics of SystemC.

3.1 Bounded Model Checking and Induction

BMC was introduced by Biere et al. in [30] and gained popularity very fast. For a LTL formula φ the basic idea of BMC is to search for counter-examples to φ in executions of the system whose length is bounded by k time steps. More formally, this can be expressed as:

$$BMC^k = I(s_0) \wedge \bigwedge_{i=0}^{k-1} T(s_i, s_{i+1}) \wedge \neg\varphi^k$$

where $I(s_0)$ denotes the predicate for the initial states, T denotes the transition relation and $\neg\varphi^k$ constraints that the property φ is violated by an execution of length k. In case of simple safety properties of the form AGp where p is a propositional formula, the violation of the property reduces to $\bigvee_{i=0}^{k} \neg p_i$, where P_i is the propositional formula p at time step i. The overall problem formulation is then transformed into an instance of *Boolean Satisfiability Problem* (SAT). If this instance is satisfiable a counter-example of length k has been found. Usually, BMC is applied by iteratively increasing k until a counter-example for the property has been found or the resources are exceeded. One of the possibilities to make BMC complete, i.e. to prove a property, is to apply induction-based methods as proposed in [31, 32]. For verifying safety properties the basic idea is to show that, if p holds after k time steps, then it must also hold after the $(k + 1)$th step. For completeness, a constraint requiring the states of an execution path to be unique has to be added.

The C Bounded Model Checker (CBMC) [33] is an implementation of BMC for C programs applying a loop unwinding technique. The execution of the program is bounded by unwinding each loop to the given bound. The unwound program is transformed into *single assignment form*, i.e. each variable is assigned exactly once, and subsequently into a set of constraints which is then solved using a SAT/SMT solver. CBMC supports assertions and assumptions embedded in the program code. Assertions are checked for all bounded execution paths of the program that satisfy the assumptions. User-input can be modeled by means of built-in non-deterministic choice functions, e.g. *nondet_int()* returns a non-deterministically chosen value of type *int*. As can be seen from the example in Fig. 1, Line 2 assigns a non-deterministic value to a variable x. Line 3 makes an assumption that the value of x is odd. As a result, x can not be zero in all execution paths after Line 3 and the assertion on Line 0 is thus satisfied.

Fig. 1 Example of assert, assume and non-deterministic value in CBMC

```
int main(int argc , char *argv[]) {
    int x = nondet_int();
    assume(x % 2 == 1);
    assert(x != 0);
    return 0;
}
```

3.2 SystemC Basics

In the following only the essential aspects of SystemC are described. SystemC has been implemented as a C++ class library, which includes an event-driven simulation kernel. The structure of the system is described with ports and modules, whereas the behavior is described in processes which are triggered by events and communicate through channels. A process gains the *runnable* status when one or more events of its sensitivity list have been notified. The simulation kernel selects one of the runnable processes and gives this process the control. The execution of a process is non-preemptive, i.e. the kernel receives the control back if the process has finished its execution or suspends itself by calling *wait()*. SystemC offers many variants of *wait()* and *notify()* for event-based synchronization such as *wait(time)*, *wait(event)*, *event.notify()*, *event.notify(delay)*, etc.

The simulation semantics of SystemC can be summarized as follows [3]: First, the system is elaborated, i.e. instantiation of modules and binding of channels and ports is carried out. Then, there are the following steps to process:

1. *Initialization*: Processes are made runnable.
2. *Evaluation*: A runnable process is executed or resumes its execution. In case of immediate notification, a waiting process becomes runnable immediately. This step is repeated until no more processes are runnable.
3. *Update*: Updates of signals and channels are performed.
4. *Delta notification*: If there are delta notifications, the waiting processes are made runnable, and then it is continued with Step 2.
5. *Timed notification*: If there are timed notifications, the simulation time is advanced to the earliest one, the waiting processes are made runnable, and it is continued with Step 2. Otherwise the simulation is stopped.

In the remainder of the chapter, we assume that a SystemC design repeatedly receives input from the environment/user. The simpler, special case, where a design receives some inputs, processes them and then terminates, is not explicitly discussed for the sake of simplicity.

4 TLM Property Checking

This section presents the property checking approach to verify transaction and system-level properties of SystemC TLM designs. Before we give the details, first a simple but conceptually representative SystemC TLM model is discussed.

As mentioned in Sect. 1, we do not target any particular TLM library and its corresponding modeling style but rather TLM designs in a broader sense, namely with transaction-based communication and event-based synchronization. The representative SystemC model contains both essential elements of TLM: event and transaction which is basically an interface function used for inter-module or inter-process communication in the model. Moreover, this model also serves as running example throughout the rest of this chapter.

Example 1 The SystemC TLM program shown in Fig. 2 models a simple communication between an initiator, a target, and a slave module using transactions and an internal event (declared in Line 19). The example has two processes: *initiate* (Line 9) from the initiator and *increase* (Line 23) from the target. The target is connected to the initiator through a port (Line 6). The process *increase* repeatedly waits for the notification of the internal event *e* before it first decreases the variable *number*, and then initiates the *transaction inc* of the slave which increases this variable again. The event *e* will be notified with a delay of 10 ns after the *transaction activate* (Line 22) of the target is called from the process *initiate* (Line 9) through the port. This transaction will be initiated every 50 ns (because of the timed *wait* in Line 12).

The current formal tools cannot directly handle a C++ implementation of SystemC (e.g. the reference implementation) due to the extensive use of C++ object-oriented features and dynamic data structures in the kernel. Therefore, we need to use a simpler and more abstract formal model as basis for property checking. Such a formal model has to capture the execution semantics of SystemC, which is clearly non-trivial even for the simple example. Moreover, the sample points for the temporal operators have to be defined, a convenient property specification language has to be identified as well as an appropriate verification method has to be found. The answers to these questions are introduced in the following subsections. Before they are presented the overall flow of our approach is illustrated in Fig. 3. At first, the model generation is performed which basically transforms the SystemC TLM model to C and integrates an abstracted SystemC kernel (see Sects. 4.1 and 4.2). Then, the monitoring logic for a concrete TLM property is built and embedded into the resulting sequential C model. This task including the property language and mappings for the different variants of TLM properties are discussed in Sect. 4.3. Finally, the BMC-based verification method and the necessary formalization to search for property violations is detailed in Sect. 4.4.

4.1 Simplified Model of the SystemC Kernel

We present a simplified model for the SystemC kernel preserving its simulation semantics. This kernel model, consisting of a kernel state vector $\mathbb{K}$ and a scheduler, allows to transform a complex SystemC design into an equivalent but much more simple C model enabling the use of formal techniques.

A SystemC TLM design is abstracted as a triple $SCTLMD = (\mathbb{S}, \mathbb{P}, \mathbb{E})$. $\mathbb{S}$ is the state vector of the design consisting of all its variables. $\mathbb{P}$ is a set of n SC_THREAD

Fig. 2 Simple SystemC TLM program

```
class activate_if : virtual public sc_interface
  { virtual void activate() = 0; };
class slave_if : virtual public sc_interface {
  { virtual void inc(int&) = 0; };
class initiator : public sc_module {
    sc_port<activate_if> port;
    initiator(sc_module_name name)
        : sc_module(name) { SC_THREAD(initiate); }
    void initiate() {
        while (true) {
            port->activate();
            wait(50, SC_NS);
        }
    }
};
class target : public activate_if, public sc_module {
    sc_port<slave_if> port;
    int number;
    sc_event e;
    target(sc_module_name name) : sc_module(name),
      number(0) { SC_THREAD(increase); }
    void activate() { e.notify(10, SC_NS); }
    void increase() {
        while (true) {
            wait(e);
            --number;
            port->inc(number);
        }
    }
};
class slave : public slave_if, public sc_module {
    slave(sc_module_name name) : sc_module(name) { }
    void inc(int& x) { ++x; }
};
int sc_main (int argc , char *argv[]) {
    initiator initiator_inst(”Initiator”);
    target target_inst(”Target”);
    ...
    sc_start();
}
```

processes $p_1, \ldots, p_n$ synchronized by the set $\mathbb{E}$ of m events $e_1, \ldots, e_m$. Each process $p \in \mathbb{P}$ is a sequence of statements. Each statement is either a C++ statement updating the design state vector $\mathbb{S}$ or a call to *wait* or *notify* for synchronization. The semantics of *SCTLMD* is only fully defined on a kernel with its kernel state vector $\mathbb{K}$. Each call to *wait* or *notify* manipulates the kernel state vector $\mathbb{K}$, which consists of the following components:

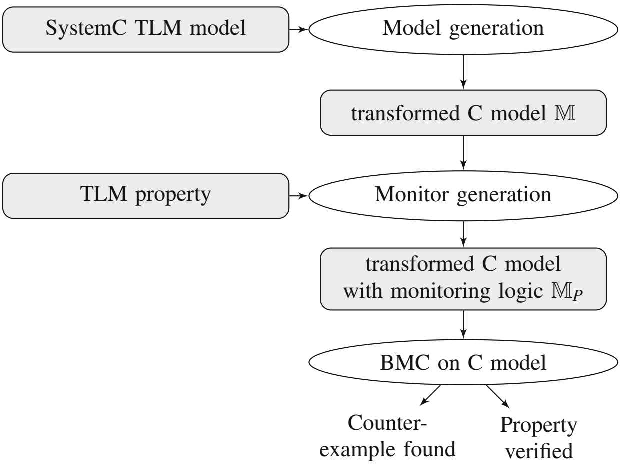


Fig. 3 Overall flow

- *globalTime*: the integer global simulation time.
- $status[p_i]$: the status of process p_i, which can be *RUNNING*, *RUNNABLE*, *WAITING* or *TERMINATED*.
- $statement[p_i]$: this indicates the next statement to be executed of p_i and basically provides the functionality of a program counter.

Algorithm 1: An interpreter for *SCTLMD*

Definition:
$\mathbb{P}_{runnable} = \{p \in \mathbb{P} \mid status[p] = RUNNABLE\}$ $\mathbb{E}_{pending} = \{e \in \mathbb{E} \mid pending[e]\}$
while $\mathbb{P}_{runnable} \neq \emptyset$ **do** `// evaluation loop`
    $p \leftarrow$ `NondetSelect(`$\mathbb{P}_{runnable}$`)`
    `Execute(`p`) // see Alg. 3`
    if $\mathbb{P}_{runnable} = \emptyset$ **then** `// delta notif. phase`
        foreach $e \in \mathbb{E}_{pending}$ **do**
            if $delay[e] = 0$ **then**
                `DoNotification(`e`) // see Alg. 2`
        end
    end
    if $\mathbb{P}_{runnable} = \emptyset$ **then** `// timed notif. phase`
        $minDelay \leftarrow \min_{e \in \mathbb{E}_{pending}}(delay[e])$
        $globalTime \leftarrow globalTime + minDelay$
        foreach $e \in \mathbb{E}_{pending}$ **do** $delay[e] \leftarrow delay[e] - minDelay$

        foreach $e \in \mathbb{E}_{pending}$ **do**
            if $delay[e] = 0$ **then**
                `DoNotification(`e`)`
        end
    end
end

Algorithm 2: DoNotification(e)

```
foreach p ∈ ℙ do
    if waiting[p, e] then
        waiting[p, e] ← false
        status[p] ← RUNNABLE
    end
end
pending[e] ← false
```

- $waiting[p_i, e_j]$: this indicates whether the execution of p_i is currently blocked by waiting for event e_j.
- $pending[e_j]$: this indicates whether e_j is pending to be notified.
- $delay[e_j]$: the integer delay of the notification of e_j.

The interpreter semantics of *SCTLMD* is defined by the interpreter in Algorithm 1. For the sake of clarity, the subroutines to interpret an individual process and to notify an event are depicted separately in Algorithm 2 and Algorithm 3, respectively. For the former task we also need the auxiliary function *next(p, stmt)* which returns the statement after the statement *stmt* of process *p*.

The preservation of the simulation semantics can be explained as follows. The loop in Algorithm 1 corresponds to the evaluation loop. In Line 2 and 3 of Algorithm 1, one of the runnable processes is selected and executed, respectively. During its execution (see Algorithm 3) the process can suspend itself (Line 4–8), or issue immediate notification (in the current delta cycle; Line 10) or delayed notification (Line 11–13). If Line 5 of Algorithm 1 is reached, it means there is no more runnable process, thus the current evaluation loop iteration is finished and the delta notification phase is entered. In this phase, all pending events with zero delay are notified. If we have at least one runnable process afterwards, the timed notification phase (Line 11–18 of Algorithm 1) is skipped and the execution continues with a new evaluation loop iteration. Otherwise, the current delta cycle is over and therefore the timed notification phase can start. Line 11–14 of Algorithm 1 advance the simulation time to the earliest pending notification and update the delays accordingly. The for-loop starting on Line 15 notifies all pending events whose delay has become zero. If those notifications make at least one process runnable, a new evaluation loop iteration starts. Otherwise, the loop condition on Line 1 fails and the simulation stops. The preservation of the simulation semantics allows us to transform the SystemC TLM design into an equivalent C model based on the simplified kernel. The generation of this C model is described in the next section.

Algorithm 3: Execute(p)

```
status[p] ← RUNNING
while statement[p] ≠ ⊥ do
    switch statement[p] do
        case wait(e)
            waiting[p,e] ← true
            statement[p] ← next(p, statement[p])
            status[p] ← WAITING
            return
        case notify(e) DoNotification(e)

        case notify(e,t)
            pending[e] ← true
            delay[e] ← min(delay[e],t)
        otherwise execute stmt[p]

    endsw
    statement[p] ← next(p, statement[p])
end
status[p] ← TERMINATED
```

4.2 *Model Generation*

We use C as our intermediate modeling language. On the one hand the transformation process is manageable and can be automated (as done in this work). On the other hand we can leverage available model checkers. The transformation into a C model consists of two steps, which also will be demonstrated for Example 1.

4.2.1 SystemC to *SCTLMD*

The first step transforms the SystemC design into the simplified form of *SCTLMD* and is divided further into two smaller steps. In the first substep, we identify the static elaborated structure of the design, that means the module hierarchy, the processes and the port bindings. With the port bindings being resolved already, every function call through a port is replaced by the call of the corresponding function of the bound module/channel instance. Afterwards the object-oriented features of SystemC/C++ are translated back into plain C. Member variables, member functions and constructors of each object instance are transformed to global variables and global functions. The result of this intermediate step for Example 1 is shown in Fig. 4. For example, the transformed code for the target module is shown in Line 8–18. The first three lines define three global variables, which were member variables of the module before (Line 18 and 19 of Fig. 2, respectively). The remaining lines show the transformed constructor *target_inst_init* and two former member functions *target_inst_activate* and *target_inst_increase*. At the beginning of *main* function, the transformed constructors are called (starting from Line 23). That corresponds to the instantiation of the modules in *sc_main*. In the second substep, all function calls in the body of the declared processes are inlined. After that all remaining function calls in process body are *notify()* and *wait()*. Thus, the declared processes contain only C++ statements and calls to *wait* or *notify* and hence form the set $\mathbb{P}$. The declared variables and events now correspond to the set $\mathbb{S}$ and $\mathbb{E}$, respectively. The SystemC design has been therefore fully transformed into *SCTLMD*.

4.2.2 Kernel Integration

The second step generates the kernel vector and the static scheduler based on the interpretation semantics described in Sect. 4.1. First, the kernel vector is added into the C model as global variables. The variables *globalTime*, $status[p_i]$ and $delay[e_j]$ are introduced as integer-valued global variables, while $waiting[p_i, e_j]$ and

```
1  void initiator_inst_init() { }
2  void initiator_inst_initiate() {
3      while (true) {
4          target_inst_activate();
5          wait(50, SC_NS);
6      }
7  }
8  int target_inst_number;
9  sc_event target_inst_e;
10 void target_inst_init() { target_inst_number = 0; }
11 void target_inst_activate() { target_inst_e.notify(10, SC_NS); }
12 void target_inst_increase() {
13     while (true) {
14         wait(target_inst_e);
15         --target_inst_number;
16         slave_inst_inc(target_inst_number);
17     }
18 }
19 void slave_inst_init() { }
20 void slave_inst_inc(int& x) { ++x; }
21
22 int main(int argc , char *argv[]) {
23     initiator_inst_init();
24     target_inst_init();
25     ...
26     sc_start();
27 }
```

Fig. 4 Result of the first substep of the "SystemC to *SCTLMD*" step

$pending[e_j]$ are declared as boolean global variables. Instead of saving and updating $statements[p_i]$ after the execution of each statement of p_i, an optimization is employed based on the observation that only statements after each potential context switch (a call of *wait()*) are relevant. For each process p_i, a label (resume point) is inserted after each call of *wait()*, and an integer-valued variable $resumePoint[p_i]$ is added to keep track of the current resume point of the process. The execution of p_i can be resumed later based on the value of $resumePoint[p_i]$ by jumping to the corresponding label. Furthermore, a counter for the number of runnable processes *runnable_count* is added. The generated scheduler for Example 1 is shown in Fig. 5. As can be seen this scheduler has the same structure as the interpreter in Fig. 1. The call of *sc_start* (Line 26 of Fig. 4) is to be replaced with this generated scheduler. In the body of the evaluation loop, non-deterministic choice, i.e. which runnable process is to be executed next, is implemented (Line 2 of Fig. 5). This non-deterministic choice allows a C model checker to explore *all interleavings* implicitly. In case the design contains no delta/timed notifications, the delta/timed notification phase is unnecessary and can be removed. The example has no delta notification but only one timed notification *e.notify(10, SC_NS)* and one timed wait *wait(50, SC_NS)* which is also modeled as timed notification. The code in Fig. 6 models the timed notification phase. Note that the event *timeout* has been introduced to implement *wait(50, SC_NS)*. Line 1–10 correspond to Line 11–14 of Algorithm 1. Line 11–26 show the implementation of the for-loop on Line 15 of Algorithm 1. Figure 7 shows the body of the process *target_inst_increase*. There is only one resume point in this process defined on Line 8. The first two lines implement the resuming of the process execution. Line 4–7 show the implementation of *wait(target_inst_e)*.

4.2.3 Limitations

The first step can handle SystemC TLM designs without dynamic process/object creation, recursion, and dynamic memory allocation. These restrictions do not

Fig. 5 Generated scheduler for the example

```
1  while (runnable_count > 0) { // evaluation loop
2      choose_one_runnable_process();
3      runnable_count−−;
4      if (initiator_inst_initiate_status == RUNNING)
5          initiator_inst_initiate();
6      if (target_inst_increase_status == RUNNING)
7          target_inst_increase();
8      if (runnable_count == 0) { // delta notification phase
9          ...
10     }
11     if (runnable_count == 0) { // timed notification phase
12         ...
13     }
14 }
```

Fig. 6 Timed notification phase

```
min_delay = 0;
if (target_inst_e_pending &&
      (min_delay == 0 || target_inst_e_delay < min_delay))
      min_delay = target_inst_e_delay;
if (timeout_pending &&
        (min_delay == 0 || timeout_delay < min_delay))
      min_delay = timeout_delay;
global_time += min_delay;
if (target_inst_e_pending) target_inst_e_delay −= min_delay;
if (timeout_pending) timeout_delay −= min_delay;
if (target_inst_e_pending && target_inst_e_delay == 0) {
      if (target_inst_increase_waiting_target_inst_e) {
            target_inst_increase_waiting_target_inst_e = false;
            target_inst_increase_status = RUNNABLE;
            runnable_count++;
      }
      target_inst_e_pending = false;
}
if (timeout_pending && timeout_delay == 0) {
      if (initiator_inst_initiate_waiting_timeout) {
            initiator_inst_initiate_waiting_timeout = false;
            initiator_inst_initiate_status = RUNNABLE;
            runnable_count++;
      }
      timeout_pending = false;
}
```

Fig. 7 The body of *target_inst_increase*

```
if (target_inst_increase_resume_point==1)
      goto target_inst_increase_resume_point_1;
while (true) {
      target_inst_increase_waiting_target_inst_e = true;
      target_inst_increase_status = WAITING;
      target_inst_increase_resume_point = 1;
      goto target_inst_increase_end;
      target_inst_increase_resume_point_1: ;
      −−target_inst_number;
      ++target_inst_number;
}
target_inst_increase_status = TERMINATED;
target_inst_increase_end: ;
```

severely limit the practical use of the proposed approach because still a very wide range of SystemC TLM models does not use the mentioned elements. Regarding the supported language constructs, currently only a subset of SystemC can be transformed which includes the core modeling components (modules, ports, interfaces, and channels) and the event-based synchronization constructs. Many other constructs are built on these fundamentals and including those in the supported subset is only an implementation issue.

The equivalence of the generated C model and the SystemC design depends largely on the kernel abstraction. Unfortunately, formally proving that any implementation of the SystemC kernel (be it our abstraction or the reference implementation) is correct with respect to the informal English specification [3] is especially hard. However, our generated C model is also executable. This has enabled simulation-based equivalence checking between the generated and the original models. This check has been employed extensively and the obtained positive results have greatly increased the confidence in the correctness of the abstraction.

In the remainder of this chapter we denote the generated C model as $\mathbb{M}$. In the next section we present our property language and how to monitor properties in $\mathbb{M}$.

4.3 Property Language and Monitor Generation

For property specification we use PSL [26] which initially was not designed for property specification at high level of abstraction. In [25] additional primitives have been introduced—coming from the software world—which are well suited for TLM property specification. Besides the variables in the design we use the following:

- *func_name:**entry***—start of a function/transaction
- *func_name:**exit***—end of a function/transaction
- *event_name:**notified***—notification of an event
- *func_name:number*—return value in case number = 0 and parameters of a function/transaction otherwise.

It is left to define the temporal sampling rate as well as the supported temporal operators. As **default temporal resolution** we sample at all *system events*, which is either the start or the end of any transaction or the notification of any event. The construct *default clock*[2] of PSL can be used to change the temporal resolution, e.g. to sample only at notification of a certain event. As temporal operators we allow ***always*** and ***next***.

The semantics of a property can be defined formally with respect to an execution trace of the SystemC TLM design. Let $se_0 se_1 \ldots se_n$ be the sequence of system events that occurred during an execution. We use $\mathbb{S}(se_i)$ to denote the state vector of the design sampled at se_i. For $\tau = \mathbb{S}(se_0)\mathbb{S}(se_1)\ldots\mathbb{S}(se_n)$, we use $\tau \models P$ to express that a property P is satisfied by τ, $\tau(i)$ for the ith element of τ, τ_i for the subsequence starting from the ith element, and $|\tau|$ for the number of elements in τ. The formal semantics of a property P with respect to τ is defined as follows.

- $\tau \models p$ iff the atomic proposition p holds in $\tau(0)$.
- $\tau \models !P$ iff $\tau \not\models P$

[2]In the considered TLM models there are no clocks. We only use the clock expression syntax to define sampling points.

- $\tau \models P \ \&\& \ Q$ iff $(\tau \models P) \wedge (\tau \models Q)$
- $\tau \models P \,||\, Q$ iff $(\tau \models P) \wedge (\tau \models Q)$
- $\tau \models \boldsymbol{next\, P}$ iff $\tau_1 \models P$
- $\tau \models \boldsymbol{always\, P}$ iff $\forall i < |\tau| : \tau_i \models P$

In the following we discuss different useful types of properties and the generation of *monitoring logic* by means of FSMs. The task of the monitoring logic is to check whether the property holds during the execution of the design.

4.3.1 Simple Safety Properties

This type of properties concern values of variables of the TLM model at any time during the execution, e.g. the values of some certain variables should always satisfy a given constraint. Generally, this property type can be expressed by a C logical expression. To verify those properties we only need to insert assertions right after the lines of code that change the value of at least one variable involved. As an example see the property depicted at the top of Fig. 8 specified for a FIFO.

4.3.2 Transaction Properties

This type of properties can be used to reason about a transaction effect, e.g. checking whether a request or a response (both are parameters or return value of some functions) is invalid or whether a transaction is successful. Monitoring logic for these properties is created by inserting assertions before/after the body of corresponding inlined function calls. For example, the property "the memory read transaction always operates on a valid address" for a TLM bus can be formulated in a transaction property as shown in the middle of Fig. 8. Recall that *mem_read:1* refers to the first parameter describing the address of the transaction.

—- Simple safety property:
// the number of processed blocks never exceeds the number of blocks
// which have been read
always *(num_block_processed <= num_block_read)*
—- Transaction property:
// the memory read transaction always operates on a valid address
always *(0 <= mem_read:1 && mem_read:1 <= MAX_ADDR)*
—- System-level property:
// Two properties for running Example 1
always *(target_inst.activate:***exit** *−>***next** *(target_inst.e:***notified** *&&* **next** *slave_inst.inc:***entry***))*
always *(slave_inst.inc:***entry** *−>***next** *(slave_inst.inc:***exit** *&& target_inst.number == 0))*

Fig. 8 Several example properties

4.3.3 System-Level Properties

These properties focus on the order of occurrences of event notifications and transactions, e.g. a given transaction should only begin after a certain event has been notified. We implement the monitoring logic using FSMs. Each state of the FSM corresponds with one position in the order specified by the property. Code for transitions of the FSM is inserted right after event notifications, begin or end of transactions (depending on the property). The FSM also has one state indicating the violation of the property. Our assertion is that this state is never reached. As example see the lower part of Fig. 8. The first system-level property has been specified for Example 1 and states that after the transaction *activate* has finished the event is notified which causes the transaction *inc* to start. The second system-level property defines the expected value of the integer *number* of the target at the end of the transaction *inc*.

Recall that the C model $\mathbb{M}$ has been automatically generated from the SystemC TLM design. Now, the monitoring logic for P is generated and embedded automatically into $\mathbb{M}$ creating a new model $\mathbb{M}_P$. This new model includes in addition to $\mathbb{S}$ and $\mathbb{K}$ a set of new variables $\mathbb{L}$ used in the monitoring logic. The next section gives a detailed presentation on verifying the C model $\mathbb{M}_P$ using BMC.

4.4 BMC-Based Verification

First of all we explain the notion of states and how the transition relation is formed with respect to the generated C model including the assertions, i.e. $\mathbb{M}_P$. The basic idea is to view the values of the variables in $\mathbb{S} \cup \mathbb{K} \cup \mathbb{L}$ as a state s and each iteration of the evaluation loop of the scheduler as the transition relation T. Each execution of the model can be formalized as a path, which is a sequence of states $s_{[0..n]} = s_0 s_1 \ldots s_n$ satisfying the condition

$$path(s_{[0..n]}) = \bigwedge_{0 \le i < n} T(s_i, s_{i+1}).$$

Note that the path can be infinite, in that case $n = \infty$.

The property P holds in the original design, iff the *general property* “no assertions fail” holds in $\mathbb{M}_P$, which also means no assertion failure during each iteration of the evaluation loop, or in other words during each transition $T(s_i, s_{i+1})$.

Definition 1 A transition without assertion failure is called *safe* and written as $safe(s_i, s_{i+1})$.

Thus, for the property to hold, every sequence of states of an execution must satisfy as well the condition

$$allSafe(s_{[0..n]}) = \bigwedge_{0 \leq i < n} safe(s_i, s_{i+1}).$$

The need for safe transitions instead of the conventional safe states is explained as follows. The transition relation in our context is defined by the evaluation loop. Therefore, a state and its successor state are defined at the start and at the end of each iteration of the evaluation loop, respectively. When an assertion fails, the execution is immediately stopped somewhere in the middle of an iteration. We already left the last state but have not reached the next one yet. It follows that the need for safe transitions is directly implied by the way the monitoring logic is generated. On the other hand, if we want to use the notion of safe states, the monitoring logic must be modified as follows. We would need to add one more Boolean flag indicating whether the property is already found to be violated. Then, each assertion would be replaced with a piece of code, that raises the flag and jumps to the end of the evaluation loop, where the flag is asserted to be false. However, this extension makes the model and its state space bigger, thus using safe transitions is actually better.

Let I be the characteristic predicate for all initial states, which are reachable states before entering the evaluation loop—note that there can be more than one initial state because some variables are uninitialized or modeled as inputs, and thus have a non-deterministic value. Then, the BMC problem can be formulated as proving that there exists an execution path of length k, starting from an initial state, and containing unsafe transitions. This is encoded in the following formula:

$$\exists s_0 \ldots s_k.\ (I(s_0) \wedge path(s_{[0..k]}) \wedge \neg allSafe(s_{[0..k]}))$$

Now BMC checks the formula for increasing k starting from zero. In the experiments (see Sect. 6.1) we show that this already gives good results. At our level of abstraction, for a fixed value of k checking the above formula is equivalent to verify the program $\mathbb{M}_P^k$, which is $\mathbb{M}_P$ with the evaluation loop unwound k times. Now there are two possible outcomes:

1. If a trace is returned, the formula is proven to hold for that fixed value of k, and the property P is proven to be false. This trace can be easily converted to an error trace for the original TLM design: the values of the variables in $\mathbb{S}$ are extracted and mapped back to the original variables, while the values of the variables in $\mathbb{K}$ are used to derive the scheduling sequence. This error trace can then be replayed on the original design by any SystemC implementation that supports user-defined scheduling sequence.
2. Otherwise, the property holds up to k. Recall that as mentioned above a SystemC design repeatedly receives input from the environment/user. Hence, for a complete proof the property has to hold for all values of k. In principle, we do not need to check a transition more than one time, thus we can stop increasing k if it reaches the number of states. However, this becomes infeasible very fast. Hence, we devise a method using induction where we derive much better

terminating conditions. The main advantages of the induction-based method are that much better run-times are achieved and the method is complete, i.e. properties are proven not only up to a certain bound k but under all circumstances.

Back to the example, assume that we want to check the property ***always*** *(target_inst.number == 0).* The monitoring logic is generated by inserting the statement *assert(target_inst_number == 0);* after Line 9 and 10 of Fig. 7, respectively. The transition relation T corresponds to the body of the evaluation loop in Fig. 5. Checking the property up to the bound $k = 3$ means applying a C model checker to a program consisting of three repetitions of the loop body. Since the example has no inputs or uninitialized variables, there is only one initial state. The used C model checker returns a trace indicating an assertion violation. The extracted execution path shows the violating scheduling sequence *initiate, increase, increase*. The violation occurs during the third transition (i.e. the second execution of *increase*) right after Line 9 of Fig. 7 is reached. On the other hand, the last property in Fig. 8 holds which essentially states that *number* is always zero at the end of the transaction *inc*. This shows the importance of the temporal resolution used in properties.

5 Induction-Based TLM Property Checking

This section introduces the induction-based method which forms a complete and efficient approach to prove transaction and system-level properties. Traditional induction-based techniques (like e.g. [31]) addressed safety state properties in the context of circuits, whereas our general property *allSafe* for $\mathbb{M}_P$ involves transitions (pairs of states) and our level of abstraction is higher. Nevertheless, the underlying ideas give a good starting point.

First, we only need to check each transition once, thus only paths, where all states except the last one are different, need to be considered. Second, after the base case is proved (i.e. no counter-example of length up to k exists), we check two terminating conditions: the "forward" condition and the inductive step. The forward condition checks whether a path of length $(k + 1)$ starting from an initial state exists. The inductive step checks if a path with k first safe transitions and a last unsafe one exists. If no such paths exist, we can stop and conclude that the property P holds. In summary, the forward condition checks the satisfiability of

$$I(s_0) \wedge loopFree(s_{[0..k]})$$

and the inductive step checks the satisfiability of

$$loopFree(s_{[0..k]}) \wedge allSafe(s_{[0..k]}) \wedge \neg safe(s_k, s_{k+1})$$

where

$$loopFree(s_{[0..k]}) = path(s_{[0..k]}) \wedge \bigwedge_{0 \leq i < j \leq k} s_i \neq s_j.$$

Now, to make induction possible at our level of abstraction, the main challenge is the embedding of the constraints into the transformed C model. Conceptually, new variables s_0, s_1, ... are needed to capture the state s after each iteration of the evaluation loop and the constraints can then be imposed by means of assumptions.[3] For example, the constraint $loopFree(s_{[0..k]})$ can be emulated by inserting the statement assume(newState(i)); after the ith unwound iteration of the evaluation loop with

$$newState(i) = (s! = s_{i-1})\&\& \ldots \&\&(s! = s_0).$$

For a precise description of the algorithm, we use the interpretation of C programs as strings. As defined earlier, $\mathbb{M}_P^k$ is $\mathbb{M}_P$ with the evaluation loop unwound k times. Let $\mathbb{M}_P[i]$ be the code fragment of the ith unwound iteration of the evaluation loop of $\mathbb{M}_P$ and let + be the string concatenation operator. Then, we have

$$\mathbb{M}_P^k = \mathbb{M}_P[1]\mathbb{M}_P[2] \ldots \mathbb{M}_P[k] = \sum_{i=1}^{k} \mathbb{M}_P[i].$$

Additionally, we define $\mathbb{M}_{\widehat{P}}$ as the resulting program after each assertion related to P in $\mathbb{M}_P$ is substituted by an assumption. The introduced notation applies for $\mathbb{M}$ and $\mathbb{M}_{\widehat{P}}$ as well. We end up with a *high-level strengthened induction with depth* shown in Algorithm 4, which has a similar structure as the Algorithms 3 and 4 in [31], but the level of the induction differs significantly. The base case is checked in the first if-statement (Line 3). The second if-statement (Line 5) checks the forward condition. The assertion at the end of Line 5 is the unwinding assertion for the evaluation loop. The last if-statement (Line 7) is the inductive step. The first arbitrary state is emulated by the statement $s = non\_det$, which assigns non-deterministic values to the variables and thus allows the model checker to examine all possible values implicitly. The underlying C model checker is invoked to verify the assertions in the passed parameter by the call *CPROVER*, which returns *true* if no assertions are violated and *false* otherwise, in this case a violating trace can be extracted.

[3] C model checkers typically support an assumption concept, i.e. assertions are checked for all execution paths of the program that satisfy the assumptions.

Algorithm 4: High-level strengthened induction with depth for transformed SystemC TLM model including monitoring logic

$k \leftarrow$ some constant which can be greater than zero
while *true* **do**
  if *not CPROVER*$(\mathbb{M}_P^k)$ **then**
    return Trace $c_{[0..k]}$
  if $CPROVER\Big(\sum_{i=1}^{k}\big(s_{i-1} = s; \mathbb{M}[i]\ \text{assume}(\text{newState}(i));\big)$
    $\text{assert}(!(\text{runnable\_count} > 0));\Big)$ **then**
    return *true*
  if $CPROVER\Big(s = non\_det; \sum_{i=1}^{k}\big(s_{i-1} = s; \mathbb{M}_{\widehat{P}}[i]$
    $\text{assume}(\text{newState}(i));\big)\ \mathbb{M}_P[k+1]\Big)$ **then**
    return *true*
  $k \leftarrow k + 1$
end

Algorithm 4, while not depending on any particular C model checker, has one possible limiting factor in the way the constraint *loopFree* is imposed. As described above, to capture the current state after each unwound iteration of the evaluation loop, auxiliary variables need to be introduced enlarging the model size considerably before being given to the C model checker. This necessity can be eliminated efficiently in the case that CBMC is applied. We take advantage of the fact that CBMC transforms the program to verify into single static assignment form first. Basically, this transformation creates for each assignment a new version of the left-hand side variable. This newly created version then substitutes the variable in the assignment and in the subsequent program flow until another assignment to the same variable occurs. As a result, the most recent version of each state variable at the end of each iteration of the evaluation loop holds the value of the original variable at that point. Therefore we only need a simple static analysis to identify those most recent versions and impose the constraints on them directly.

As a final note, our approach can deal with nested loops, which are commonly present in our transformed models. The evaluation loop is handled explicitly as described above. The other loops must be unwound up to at least their run-time bounds before applying our approach. Those run-time bounds can be determined with the aid of unwinding assertions [33]. Also note that unbounded loops can still have a finite run-time bound due to the simulation semantics of SystemC and our transformation method. As an example, consider the infinite loop *while (true)* {*body1; wait(e); body2;*} commonly used in a SC_THREAD. It only needs to be unwound twice. In its first execution, the SC_THREAD performs *body1;* then suspends itself because of *wait(e)*. Any further execution resumes exactly after the wait statement, performs *body2;* then *body1;* and is suspended again by *wait(e)*. Figure 9 shows the transformation with two unwound iterations for a SC_THREAD named *t*.

Fig. 9 The infinite SC_THREAD loop unwound

```
if (t_resume_point==1)
      goto resume_point_1;
// begin of the first unwound iteration
body1;
{ // wait(e);
      t_waiting_e = true;
      t_status = WAITING;
      t_resume_point = 1;
      goto t_end;
      t_resume_point_1:
}
body2;
// end of the first unwound iteration
// begin of the second unwound iteration
body1;
{ // wait(e);
      t_waiting_e = true;
      t_status = WAITING;
      t_resume_point = 1;
      goto t_end;
}
body2;
// end of the second unwound iteration
t_end: ;
```

6 Experiments

The proposed approach has been implemented as the tool *SCIVER* and evaluated on different TLM designs.[4] The model checker CBMC v3.7 [33] with Boolector v1.2 [34] as the underlying SMT solver has been used to verify the generated C models. CBMC v3.7 does not fully support SMT yet, therefore we needed to switch back to use the default SAT solver in some cases. The proposed approach has been built on top of CBMC, i.e. unwinding and the transformations for induction are performed before giving the problem to CBMC. The internal slicer of CBMC is activated to remove subformulas, which are irrelevant for the properties. All experiments have been carried out on a 3 GHz AMD Opteron system with 4 GB RAM running Linux. The time limit for each run is set to 1 h.

In the first part of the experiments we present the results for our BMC-based verification approach which is denoted as *SC-BMC*. Then, in the second part we give the results for the induction-based method *SC-IND* as introduced in Sect. 5. We also have compared our approach to the approach ESST presented in [14, 24]. This is to the best of our knowledge the most recent approach and also the only one available to evaluate. We used the version of ESST presented in [14] with POR enabled. Note that [24] also presented results for applying some C model checkers directly to the generated sequential C model, but ESST outperformed these in most

[4]Our benchmarks can be found on the SCIVER Website: www.systemc-verification.org/sciver.

cases. As mentioned in Sect. 2, ESST does not support any properties besides local C assertions. Thus, we needed to add monitoring logic before giving a model to ESST.

6.1 BMC-Based Verification

6.1.1 FIFO Design

The first design is the *simple_fifo* TLM example included in the official SystemC distribution. The original design consists of a consumer module and a producer module communicating over a FIFO channel. Both modules have their own SC_THREAD. We modified the design so that the producer writes an infinite sequence of arbitrary characters into the FIFO. The SystemC model (the generated C model) has approximately 80 (150) lines of code. We considered the following properties of the FIFO (also listed in Fig. 10 in PSL syntax):

- P1: The number of elements in the FIFO never exceeds the limit.
- P2: After a *write* transaction, the FIFO is not empty.
- P3: If the FIFO is full, the next event notified is *read_event*.

```
---- for FIFO design:
P1: always (0 <= num_elements && num_elements <= max)
P2: always (write:exit -> num_elements > 0)
P3: default clock = read_event:notified || write_event:notified; always (num_elements == max
        -> next read_event:notified)
P4: default clock = read_event:notified || write_event:notified; always (read_event:notified ->
        next_e[1:10] write_event:notified)
---- for TLM-2.0 design:
P5: default clock = target.b_transport.entry;
always( (target.b_transport:1.command == TLM_READ_COMMAND)
   -> next( (target.b_transport:1.command == TLM_READ_COMMAND) -> next((target.
         b_transport:1.command != TLM_READ_COMMAND))))
P6: default clock = target.b_transport.exit;
always( target.b_transport:1.command == TLM_READ_COMMAND ->
          ( ((target.b_transport:1.data[3] << 24) | (target.b_transport:1.data[2] << 16) | (
                 target.b_transport:1.data[1] << 8) |
          target.b_transport:1.data[0] == target.b_transport:1.address) || ((target.b_transport:1.
                 data[3] == 0) && (target.b_transport:1.data[2] == 0) && (target.b_transport
                 :1.data[1] == 0) && (target.b_transport:1.data[0] == 0)) )
---- for JPEG encoder:
P7: always (mem_write:0 && mem_read:0)
P8: always (write:1 <= 15)
P9: default clock = read_block:exit || zigzag_scan:exit || rle_encode:exit; always (zigzag_scan:
        exit -> next[3] zigzag_scan:exit)
```

Fig. 10 Properties used in the experiments

- P4: After a notification of *read_event*, the next 10 (the FIFO size) notifications includes at least one notification of *write_event*.

The design malfunctioned as soon as we tried to connect more consumers or producers to the FIFO channel. We fixed the implementation and proved the properties on the corrected design. Since the BMC-based method from Sect. 4.4 is incomplete, the properties can only be verified for a fixed number of inputs. The results are shown in Table 1. Each sub-column gives the run-times required to verify each property for the bounded input of 48, 64 and 80 arbitrary characters, respectively. As can be seen the run-times increase with the number of characters. Moreover, adding another consumer or producer also results in higher run-times. Note that we did not apply ESST in this experiment. The comparison to ESST will be made in the next section where the properties must be proven on the unbounded design.

Table 2 shows the run-times needed to disprove P1 on the original FIFO with 2 consumers and 1 producer (fifo_2c_1p_orig), and with 1 consumer and 2 producers (fifo_1c_2p_orig), respectively. In both cases, 48 characters are written into the FIFO. The results shown in the second and third column are obtained by using our approach and ESST, respectively. As can be seen SC-BMC outperforms ESST clearly.

6.1.2 TLM-2.0 Design

The approach described in this chapter applies for general TLM models. We demonstrate the potential of extending the approach for the TLM-2.0 standard on a design consisting of two initiators connected to one target by TLM-2.0 sockets. All modules are implemented as *loosely timed* models using the TLM-2.0 interfaces.

Table 1 Results of SC-BMC for corrected FIFO

	1 consumer + 1 producer		
	48 chars (s)	64 chars (s)	80 chars (s)
P1	13.55	27.22	36.85
P2	13.08	25.98	33.83
P3	20.88	42.86	55.44
P4	14.62	29.54	38.95
	2 consumers + 1 producer		
P1	115.15	231.83	314.08
P2	104.43	221.90	286.01
P3	159.23	329.87	441.93
P4	118.58	252.91	327.81
	1 consumer + 2 producers		
P1	223.32	495.90	637.32
P2	239.82	506.41	602.09
P3	317.62	701.52	844.81
P4	242.39	525.57	646.36

Table 2 Results of SC-BMC for disproving properties

	SC-BMC (s)	ESST (s)
fifo_2c_1p_orig + P1	2.86	219.09
fifo_1c_2p_orig + P1	3.87	305.90
tlm2_bug + P5	0.62	Not applicable
tlm2_bug + P6	48.60	Not applicable

The first (second) initiator writes to (reads from) the memory of the target by using the blocking interface *b_transport* of the target. This interface is a part of the TLM-2.0 standard accepting a payload and a delay as parameters. A payload in TLM-2.0 standard carries all information regarding a transaction, i.e. command, data, address, etc. For example, each payload sent by the first initiator contains among the other things: a *TLM_WRITE* command, a non-deterministic unsigned integer multiple of four as address, and the address value as data. Note that the data field of the payload in TLM-2.0 is modeled as an unsigned char pointer allowing fast data transport. Hence, in the example the data field actually points to the first of four bytes which combined together represent the address value. The delay parameter is related to the concept of *temporal decoupling* which basically allows a process to have its own local time. In our design, each initiator keep a variable describing how far it is ahead of the simulation time. This variable is given to *b_transport* of the target as the delay parameter, thus indicating the future point in time that the transaction actually starts. The target then increases this delay parameter by an amount of time modeling the transaction duration. There exists a limit for the amount of time each initiator can be ahead of the simulation time. If this limit is exceeded, the corresponding initiator explicitly calls *wait* to allow the simulation time to catch up. Two properties have been formulated for the read transaction of the design:

- P5: At most two consecutive read transactions can occur.
- P6: At the end of a read transaction, the address and the data of the payload should be hold the same value or the data should be equal to zero (i.e. the address has not been written to).

Both are also listed in the middle of Fig. 10 in PSL syntax. The first property reasons about the synchronization of the design, while the second checks the functionality of the target. The run-times needed to disprove P5 and P6 on a buggy design (tlm2_bug) with incorrect time limit for explicit synchronization and incorrect implementation of memory write are shown in Table 2. As can be seen, SC-BMC was able to find a counter-example fairly quickly for each property. ESST cannot be applied on this design mainly due to its missing capability in handling the pointers in the TLM-2.0 payload.

6.2 Induction-Based Verification

6.2.1 FIFO Design and TLM-2.0 Design

We applied the proposed induction-based method to the FIFO design and the correct variant of the TLM-2.0 design discussed in the previous section. The results are shown in Table 3 under the column “SC-IND”. For the FIFO design, significant improvements over the BMC-based method with respect to run-time can be observed. SC-IND has proven each of the four considered properties in under 10 s, while the time in seconds needed by SC-BMC is a three-digit number in many cases (see Table 1). Furthermore, using induction, the proofs are *complete*, i.e. the properties are verified for unbounded input where arbitrary characters are repeatedly written into the FIFO. Our results can be compared to the results of using ESST which are also presented in Table 3. As can be seen, both P2 and P3 seem to be easy to prove for both approaches. For proving P1 and P4, the performance of our approach is much better. For the TLM-2.0 design, property P5 has been proven quickly using our approach and the proof is also complete. SC-IND failed to prove P6 within the time limit which seems to be very hard considering the large state space that includes the 512 bytes memory of the target.

Table 3 Results of SC-IND for proving properties with comparison to ESST

Benchmark	SC-IND (s)	ESST (s)
fifo_1c_1p_corr + P1	0.20	119.00
fifo_1c_1p_corr + P2	0.18	0.28
fifo_1c_1p_corr + P3	2.70	0.35
fifo_1c_1p_corr + P4	2.11	351.06
fifo_2c_1p_corr + P1	1.04	120.21
fifo_2c_1p_corr + P2	0.59	0.48
fifo_2c_1p_corr + P3	8.99	0.49
fifo_2c_1p_corr + P4	3.92	358.84
fifo_1c_2p_corr + P1	0.20	126.09
fifo_1c_2p_corr + P2	0.47	0.41
fifo_1c_2p_corr + P3	9.61	0.50
fifo_1c_2p_corr + P4	6.53	372.23
tlm2 + P5	3.38	Not applicable
tlm2 + P6	Timeout	Not applicable
jpeg + P7	4.95	Timeout
jpeg + P8	5.20	Timeout
jpeg + P9	453.52	Timeout

6.2.2 JPEG Encoder

As another SystemC model we considered our TLM implementation of a part of a JPEG encoder consisting of seven modules shown in Fig. 11: a simple bus, two memory slaves, a reader, a zig-zag scanner, a run-length encoder, and a controller. The SystemC model (the generated C model) has approximately 200 (450) lines of code. The input of the design are quantized DCT 8×8 pixel blocks and the output are run-length encoded sequences. The reader loads a 8×8 block of integers and writes it row for row into the first 64 cells of the memory. The scanner reads the block from the memory, rearranges it using zig-zag pattern and writes the result into the next 64 memory cells. The encoder processes data stored in those cells and writes the encoded sequence directly to output. The reader, the scanner and the encoder are controlled by the SC_THREADs *read*, *scan* and *encode* in the controller, respectively. The synchronization between the threads is as follows. After each 8×8 block is read into memory, the *read* process must wait for the *scan* process to finish the zig-zag scan on this current block, after that a new block can be read. The *encode* process can only be active after the completion of the scan on the current block. The *scan* process becomes runnable after each block is loaded into memory, but not before the RLE encoding process on the last block is finished. All memory accesses are through the *mem_read* and *mem_write* transaction of the bus. The bus decodes addresses and forwards transactions to corresponding memory slaves.

We successfully verified the following properties using induction:

- P7: All memory accesses are successful, i.e. *mem_read* and *mem_write* of the bus always return *true*.
- P8: While encoding the 8×8 block, if 16 consecutive zeros are encountered, a pair (15, 0) should be written to the output, instead of waiting for a non-zero value. Therefore, the first parameter of the *write* method of the output module should always be less than or equal to 15.

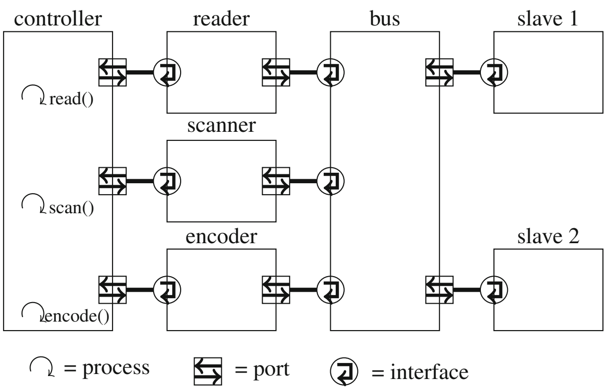


Fig. 11 A part of a JPEG encoder

Table 4 Comparison on *chain* benchmark

m (#modules)	SC-IND (s)	ESST (s)	HP w/o sched. (s)	HP w. sched. (s)
5	0.04	0.04	1.40	1.80
9	0.07	0.09	2.08	7.28
13	0.10	0.16	11.71	118.78
17	0.15	0.50	166.73	2443.77
19	0.18	0.75	686.93	Not completed
21	0.22	0.34	3201.33	Not completed

- P9: The synchronization between the threads implies, that after the completion of each scan transaction, the next scan should always follow after two other transactions.

The properties are also listed in the bottom part of Fig. 10 in PSL syntax. The results are given in Table 3. Again, the efficiency of the induction-based method can be observed and the proofs are *complete*, i.e. the properties are proven for any number of input blocks and arbitrary blocks' contents. ESST failed to prove any of the properties within the time limit.

6.2.3 Chain Benchmark

The last design considered in the chapter is the *chain* benchmark presented in [20]. This benchmark consists of a chain of *m* modules communicating through transactions. Each module has a SC_THREAD, which waits for an internal event before initiating a transaction with the next module. This transaction notifies the internal event of the next module, so that this module can start a transaction with the after next, and so on, until the last module completes its transaction. No "real" property is checked in the benchmark, instead the whole model state space is explored. The results are shown in Table 4. The first column gives the number of modules/processes in the chain. The column "SC-IND" provides our results. The results of ESST are presented in the next column, while the results from [20] using the encoding without and with a non-preemptive scheduler, respectively[5] can be seen in the last two columns. The results show that our approach and ESST are comparable on this benchmark. Both can handle a large number of processes and scales much better than [20]. Also note that the encoding with a scheduler is closer to our approach since both implement the same non-preemptive semantics.

[5]Their experiments were done on a 2 GHz AMD Opteron system with 4 GB RAM running Linux.

6.3 Summary of Experimental Results

The experiments have shown that the proposed approach is very promising. For buggy designs, SC-BMC has found counter-examples very quickly. Furthermore, SC-IND was able to prove important TLM properties efficiently in most cases. Our approach has also outperformed the recent and promising approach ESST for the considered designs. However, we expect our approach and ESST to be complementary. For example, designs with excessive number of scheduling sequences are potentially easier for ESST since it has incorporated POR techniques.

Although induction is complete, the induction depth k can be very large and thus infeasible for more complex designs. In this case we need to apply invariants to the first arbitrary state of the induction step. These invariants can be derived either manually or automatically. But this is beyond the scope of this chapter and hence left for future work.

7 Conclusions

We have presented an efficient property checking approach for SystemC TLM designs which are used as initial models when developing new cyber components of a CPS. The approach consists of three steps: the fully automated transformation of SystemC to C, the generation and embedding of monitoring logic for a TLM property, and the verification of the transformed C models. For the verification task, a BMC formulation over the evaluation loop of the scheduler has been developed. Furthermore, we improved the BMC-based technique with respect to efficiency and completeness by performing induction at the level of C programs. The experiments show that complete proofs of important TLM properties can be carried out efficiently. The large state spaces of SystemC designs, which also consist of all possible inputs and interleavings, are fully explored by our approach.

For future work we would like to investigate light-weight static POR techniques, the automatic derivation of invariants to reduce the induction depth and the use of abstraction. In addition, we also want to extend our approach to handle more TLM-2.0 constructs.

References

1. Große D, Le HM, Drechsler R (2010) Proving transaction and system-level properties of untimed SystemC TLM designs. In: ACM and IEEE international conference on formal methods and models for codesign, pp 113–122
2. Bailey B, Martin G, Piziali A (2007) ESL design and verification: a prescription for electronic system level methodology. Morgan Kaufmann, Amsterdam
3. IEEE Standard for Standard SystemC Language Reference Manual (2012) IEEE Std. 1666

4. Pasricha S (2002) Transaction level modeling of SoC with SystemC 2.0. In: SNUG
5. Cai L, Gajski D (2003) Transaction level modeling: an overview. In: IEEE/ACM/IFIP international conference on hardware/software codesign and system synthesis, pp 19–24
6. Ghenassia F (2006) Transaction-level modeling with SystemC: TLM concepts and applications for embedded systems. Springer, US
7. Vardi MY (2007) Formal techniques for SystemC verification. In: Design Automation Conference, pp 188–192
8. Dahan A, Geist D, Gluhovsky L, Pidan D, Shapir G, Wolfsthal Y, Benalycherif L, Kamdem R, Lahbib Y (2005) Combining system level modeling with assertion based verification. In: ISQED, pp 310–315
9. Ecker W, Esen V, Hull M, Steininger T, Velten M (2006) Requirements and concepts for transaction level assertions. In: International conference on computer design, pp 286–293
10. Bombieri N, Fummi F, Pravadelli G (2007) Incremental ABV for functional validation of TL-to-RTL design refinement. In: Design, automation and test in Europe, pp 882–887
11. Ferro L, Pierre L (2009) ISIS: runtime verification of TLM platforms. In: Forum on specification and design languages, pp 1–6
12. Tabakov D, Vardi MY (2010) Monitoring temporal SystemC properties. In: ACM and IEEE international conference on formal methods and models for codesign, pp 123–132
13. Ferro L, Pierre L, Amor ZBH, Lachaize J, Lefftz V (2011) Runtime verification of typical requirements for a space critical SoC platform. In: FMICS, pp 21–36
14. Cimatti A, Narasamdya I, Roveri M (2011) Boosting lazy abstraction for SystemC with partial order reduction. In: Tools and algorithms for the construction and analysis of systems, pp 341–356
15. Habibi A, Tahar S (2005) Design for verification of SystemC transaction level models. In: Design, automation and test in Europe, pp 560–565
16. Moy M, Maraninchi F, Maillet-Contoz L (2006) LusSy: an open tool for the analysis of systems-on-a-chip at the transaction level. In: Design automation for embedded systems, pp 73–104
17. Karlsson D, Eles P, Peng Z (2006) Formal verification of SystemC designs using a petri-net based representation. In: Design, automation and test in Europe, pp 1228–1233
18. Niemann B, Haubelt C (2006) Formalizing TLM with communicating state machines. In: Forum on specification and design languages, pp 285–293
19. Traulsen C, Cornet J, Moy M, Maraninchi F (2007) A SystemC/TLM semantics in Promela and its possible applications. In: SPIN, pp 204–222
20. Helmstetter C, Ponsini O (2008) A comparison of two SystemC/TLM semantics for formal verification. In: ACM and IEEE international conference on formal methods and models for codesign, pp 59–68
21. Herber P, Fellmuth J, Glesner S (2008) Model checking SystemC designs using timed automata. In: IEEE/ACM/IFIP international conference on hardware/software codesign and system synthesis, pp 131–136
22. Kundu S, Ganai M, Gupta R (2008) Partial order reduction for scalable testing of SystemC TLM designs. In: Design automation conference, pp 936–941
23. Blanc N, Kroening D (2008) Race analysis for SystemC using model checking. In: International conference on CAD, pp 356–363
24. Cimatti A, Micheli A, Narasamdya I, Roveri M (2010) Verifying SystemC: a software model checking approach. In: International conference on formal methods in CAD, pp 51–60
25. Tabakov D, Vardi M, Kamhi G, Singerman E (2008) A temporal language for SystemC. In: International conference on formal methods in CAD, pp 1–9
26. Accellera Property Specification Language Reference Manual (2005) version 1.1. http://www.pslsugar.org
27. Chou C-N, Ho Y-S, Hsieh C, Huang C-Y (2012) Symbolic model checking on SystemC designs. In: Design automation conference, pp 327–333

28. Le HM, Große D, Herdt V, Drechsler R (2013) Verifying SystemC using an intermediate verification language and symbolic simulation. In: Design automation conference, pp 116:1–116:6
29. Herdt V, Le HM, Drechsler R (2015) Verifying SystemC using stateful symbolic simulation. In: Design automation conference, pp 49:1–49:6
30. Biere A, Cimatti A, Clarke EM, Zhu Y (1999) Symbolic model checking without BDDs. In: Tools and algorithms for the construction and analysis of systems, pp 193–207
31. Sheeran M, Singh S, Stålmarck G (2000) Checking safety properties using induction and a SAT-solver. In: International conference on formal methods in CAD, pp 108–125
32. Bjesse P, Claessen K (2000) SAT-based verification without state space traversal. In: International conference on formal methods in CAD, pp 372–389
33. Clarke EM, Kroening D, Lerda F (2004) A tool for checking ANSI-C programs. In: Tools and algorithms for the construction and analysis of systems, pp 168–176
34. Brummayer R, Biere A (2009) Boolector: an efficient SMT solver for bit-vectors and arrays. In: Tools and algorithms for the construction and analysis of systems, pp 174–177

Evaluation Model for Assessment of Cyber-Physical Production Systems

Michael Weyrich, Matthias Klein, Jan-Philipp Schmidt, Nasser Jazdi, Kurt D. Bettenhausen, Frank Buschmann, Carolin Rubner, Michael Pirker and Kai Wurm

1 Introduction

Cyber-physical systems received a lot of attention for applications in various fields such as home automation, appliances, medical devices, consumer products etc. In this study, we focus on the domain of production and manufacturing: How could

M. Weyrich (✉) · M. Klein · J.-P. Schmidt · N. Jazdi
Institute of Industrial Automation and Software Engineering, University of Stuttgart, Pfaffenwaldring 47, 70569 Stuttgart, Germany
e-mail: michael.weyrich@ias.uni-stuttgart.de

M. Klein
e-mail: matthias.klein@ias.uni-stuttgart.de

J.-P. Schmidt
e-mail: jan-philipp.schmidt@ias.uni-stuttgart.de

N. Jazdi
e-mail: nasser.jazdi@ias.uni-stuttgart.de

K.D. Bettenhausen
Siemens Corporation, 755 College Road East, Princeton, NJ 08540, USA
e-mail: kurt.bettenhausen@siemens.com

F. Buschmann · M. Pirker · K. Wurm
Siemens AG, Otto-Hahn-Ring 6, 81739 Munich, Germany
e-mail: frank.buschmann@siemens.com

M. Pirker
e-mail: michael.pirker@siemens.com

K. Wurm
e-mail: kai.wurm@siemens.com

C. Rubner
Siemens AG, Günther-Scharowsky-Str. 1, 91058 Erlangen, Germany
e-mail: carolin.rubner@siemens.com

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_7

the future cyber-physical production system be evaluated and assessed and what distinguishes it from conventional production systems? To answer these questions, assessment aspects of conventional automation systems have to be summed up and placed in relation to the new ways and the impact of future systems technologies entailing "cyber abilities". This work is by nature theoretic and brings together literature reviews, interviews with experts, brainstorming ideas, analysis, and scenario analysis techniques.

The approach chosen is based on literature review and brainstorming analysis which were conducted to condense the driving forces and thereby enabling the definition of an evaluation model.

Cyber-physical systems (CPS) are defined as "integrations of computation with physical processes" [30]. The combination of physical processes with intelligent computer systems leads to new applications in various domains. In this paper, cyber-physical systems are considered in the context of production technologies, which leads to the term cyber-physical production systems. In the past few years, a lot of research has been conducted on technologies aiming for value-adds. There are multiple trends in information technology that will affect all areas of manufacturing. However, since these technologies are yet to be used by the industry on a large scale, it is difficult to estimate the impact of cyber-physical production systems. For this reason, an evaluation model was developed to assess typical characteristics of cyber-physical production systems. Based on a survey of literature on various cyber-physical system technologies, methods and tools, value-adds and their driving patterns have been identified. In this paper, the research approach to the evaluation model is described as well as the definition of the evaluation model itself.

1.1 Motivation

In the past, press articles on "Smart factories", "Industrial Internet", or "digitalization of production" have stimulated the imagination towards the impact of disruptive innovation in production. At present, there are a number of open questions about the impact cyber-physical production systems might have. What are the areas of application which are affected by cyber-physical production systems? Potentially a lot of industrial application fields will be affected by the change coming with these new technologies. However, what are these technologies about and what impact will these have on manufacturers and supply chains?

The following list gives a first impression of the future "could be" situation in some of the application fields of cyber-physical production systems:

- The automation equipment industry has reached a high level of maturity with products such as programmable logic controllers, motion controllers, field bus systems, drive technologies etc. However, a lot of manpower is consumed during engineering and set-up. Upcoming technologies like Plug and Play or

Self-X are very much needed to improve the industrial application. This raises many questions: What will new automation products look like? Will networks of microcontrollers act as decentralized control structures in the future? Would a wireless fieldbus add value? Will the Internet of Things (IoT) be the game-changer in the manufacturing industry?

- Manufacturing machinery entails a huge variety of application segments and manufacturing concepts. Improving system usability, programming, and simulation have been investigated for a long time. Likewise, research to enable and improve flexibility and changeability has been ongoing for decades, with questions such as: How can production lines become more dynamic, highly flexible and more efficient at the same time? Are there any generic approaches for production management across different industry domains? Are there any new approaches towards diagnosis and maintenance?
- Logistic is affected and could benefit strongly from new concepts. How would new products for storage or goods marking and tracking trigger a change? Will autonomous vehicles and service robots start a revolution in intra-logistics?
- The fields of IT and software for ERP, MES as well as PLM are growing. How could isolated management systems in production be merged? Is the vision of a "single source of truth" affordable? Could future IT systems support decision-making in practice? How are the enormous expenditures for IT justified?

All these upcoming technology trends basically affect the whole value-adding network of industrial production.

Are cyber-physical production systems a game changer?

The concepts of smart devices which combine software, electronics and hardware have been identified as a major driving force to change industries [37]. These technologies should be utilized for industrial production which requires to be more efficient and agile to ensure economic prosperity in high wage countries.

At present, conventional information technology is used in manufacturing automation. Research is targeting new methods and tools for a smart and intelligent production of the future. It can be expected that "the Internet of Things", "smart connected products" and other cutting-edge technology ideas are going to advance manufacturing automation and industrial production.

But, what exactly is the leverage of these new technologies? Will new technologies, such as networked production equipment gradually improve efficiency or will it be the start of a revolution? Figure 1 explains the potential impact of information and communication technology. The picture introduces so-called "cyber abilities" which are synonym for new information technologies applied to industrial production.

As a consequence, the state-of-the-art in industrial production, utilizing conventional IT (Sect. 1), is enhanced towards two potential future areas:

One is the potential to optimize the manufacturing, assembly or maintenance in a conventional production approach (Sect. 2). This would mean that the organization of production remains as per today but is gradually improved by IT. However, the other area (Sect. 3) stands for a completely new way of work based on a disruptive

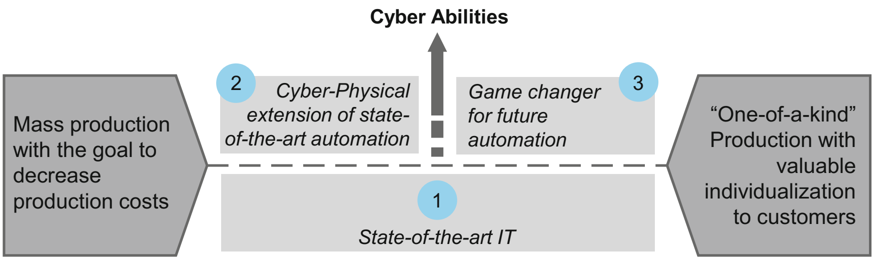


Fig. 1 The impact of cyber technology: mass production versus individual production

innovation and game changing concept. This can be referred to as "Industrie 4.0" (see: www.hightech-strategy.de). Cyber-physical production systems would act as a game changer and trigger a revolution. Thereby the manufacturing industry is hoping for an information technological breakthrough to boost industrial production in western high wage countries.

Obviously the "cyber abilities" are seen as enablers for the shift in the field of economic mass individualization for securing the innovation lead in manufacturing in developed countries, in contrast to mass production at minimal cost. The scenario envisions a highly flexible mass production, which can produce even lot size one of product by means of automated manufacturing in a highly efficient way.

2 A First Analysis of Value-Adds

What value-adds could shape the future of manufacturing automation?

A "digital production of one-of-a-kind products" or mass production accommodating endless variants is the vision from the production perspective [23]. Figure 2 presents a word cloud of production value-adds and their driving cyber technologies which is based on our research [22, 51] as well as the *acatech* report "Recommendations for implementing the strategic initiative "Industrie 4.0"" [24].

With regard to the enabling technologies, cyber-physical production systems seem to be inspired by the idea of Internet of Things (IoT) in the sense of a dynamic orchestration of intelligent units with a decentralized control [13]. This would certainly lead to changes, especially in topology and modularity of wireless networked smart units. Simulation technology also enables the analysis of complex systems.

And, there is the aspect of knowledge processing and reasoning in the sense of artificial intelligence which is gradually becoming a reality. Figure 2 shows future characteristics and capabilities of cyber-physical production systems which contribute to an improvement of various parameters and enable new business models.

The value-adds presented above are very generic and seem to reflect the various research ambitions in manufacturing automation and IT of the past decades.

Fig. 2 Typical generic value-add claim of cyber technologies and “Industrie 4.0”

Once a more detailed text analysis is conducted, a set of beneficial trends can be derived. This analysis has been undertaken to understand the value-adds of production industries, listing the value claims, the potential benefits of the technologies and the potential stakeholders. The sources of the analysis are based on a significant study in which many leading experts from the industry participated. As a result the tables below condense a solid list of value-adds as well as projections of beliefs and desires about the future of automated manufacturing.

Table 1 was condensed based on [24]. As it can be observed, the benefit of intelligent management of complexity of networked manufacturing is key and prerequisite for future manufacturing. Production planning and scheduling, logistics and maintenance aspects become more and more difficult to overlook by manufacturing managers or operators. This growing complexity in manufacturing in strongly interdependent networks leads to the desire for intelligent support capabilities [47].

Further, in cyber technology, it is assumed that the complexity of the tasks can be reduced [48] by allocating the tasks to various manufacturing units. Having the tasks distributed over decentralized manufacturing units promises optimal support while allowing for a high degree of flexibility. In particular, the orchestration of very dynamic manufacturing systems (e.g. production of individual products including complex supply chains) is desired for manageability. The reduction of complexity of tasks is one of the major steps towards the smart factory using cyber technology. Therefore, the orchestration of networked manufacturing modules to

Table 1 Overview on general value-adds of cyber technology and stakeholders

Value-add	Stakeholder	Benefit
Intelligent management of complexity of networked manufacturing	Production managers	• Intelligent support capabilities to overlook the production and logistics
Real-time availability of data along with information processing	Production managers and operators	• Decision making and intelligent reasoning
		• Provision of value-added services in future
Orchestration of networked manufacturing modules to reduce task complexity	Production operator	• Easy commissioning of automated manufacturing systems

reduce task complexity is goal in the sense of easy commissioning of automated manufacturing systems.

Real-time availability of data along with information processing to assist decision making in production management is certainly a very important value-add for any decision maker. With higher levels of cyber technologies, more and more structured data and knowledge is stored and can be processed and used. This analysis of data allows intelligent reasoning and the provision of value-added services in future.

Table 2 contains additional value-adds articulated in use cases analysis ([24]-German Appendix). They depict value-adds which are to be seen as particularly novel and can be described with respect to various stakeholders.

A large set of additional value-adds or variations of the value-adds above can be found by searching the available literature. Some of these examples are for instance, the optimization of value-adding network to lower costs and increased flexibility (i.e. change of suppliers). Another interesting aspect is the reliability of the supply-chain as well as multiple goal optimisation e.g. to reduce energy consumption and CO_2 pollution. Furthermore there is the topic of end-to-end engineering across the entire life-cycle which can be described in detail focusing on values (i.e. automatic rescheduling after supply-chain disturbances) and beneficial approaches for reliability.

3 State-of-the-Art

A large variety of different technologies are presently contributing to concepts of cyber-physical production systems. But, which technologies are generic and address multiple industry segments relevant to an evaluation model?

A lot of research has been undertaken with regard to the assessment of automation technology of the passed decades. This chapter will give an overview of this development and it will conclude with technological trends. The state-of-the-art

Table 2 Particular value-adds of cyber technology and stakeholders

Value-add	Stakeholder	Benefit
Ad hoc rescheduling in the sense of reconfiguration, flexibility and agility, i.e. Plug and Play of cyber-physical production systems	Production manager and operators	• Shorter planning and configuration times
		• Lower qualification level for engineers
		• Efficient usage of available manufacturing capacities
	Consumer	• Availability of highly individualized products at low cost
Capability of manufacturing multiple product variants, i.e. manufacturing one-off items with lot size one. (This means very individual/specific customer requirements could be met)	Production manager and operators	• Competitive advantage in global market
		• Larger product portfolio with small investment
	Consumer	• Availability of highly individualized products with high quality, low cost and short production times
Fast maintenance, diagnosis and easy operation of manufacturing systems, e.g. by means of remote services for diagnosis and operation of machines	Operator/service engineer	• Faster and easier maintenance
		• Less errors using the system
		• Lower qualification level for users
	Equipment manufacturer	• Better service through data availability by integrating data from equal components
		• Back propagation of experience into new designs
	Production manager	• Higher availability of the production
		• More reliable supply chains
	Consumer	• Increased product quality

can be grouped into three clusters. The clusters are based on research conducted by different groups addressing the topics from their individual perspectives.

Research on cyber-physical systems
Cyber physical systems were suggested by Edward Lee [28, 29] which prompted impulses of new generation possible designs of automated systems to produce individual products. This led to the present research which is focused on new business models on the basis of CPS. Cyber-physical production systems are by definition split into two levels: the physical and the cyber level. The physical indicators describe the characteristics of physical objects, e.g. hardware of automation systems which is, as mentioned, well covered by performance indicators. The "cyber" aspect might have an impact and enhance the performance indicators.

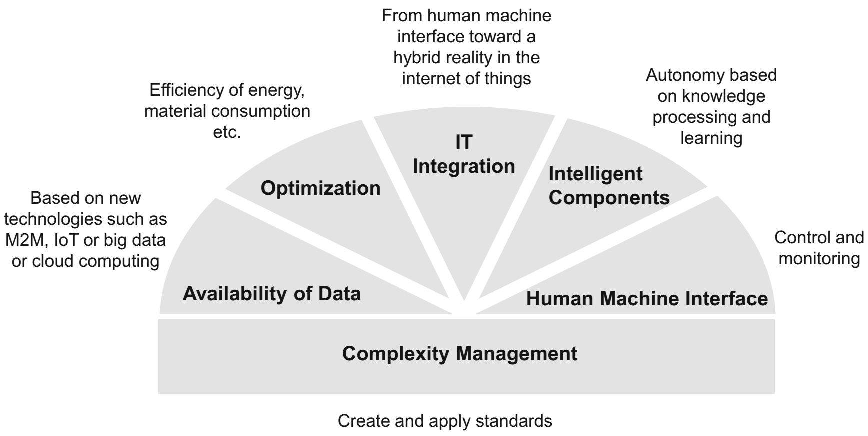


Fig. 3 Technological trends of today (see also [54])

Typical trend technologies which are often mentioned within relation to cyber-physical production systems are depicted in Fig. 3.

The availability of data is a trend which is based on technological breakthrough in processing large quantities of data. Cloud computing can improve the processing of data and information. The aspect of optimization could for instance mean, that a production system produces the products with a high energy consumption only when the costs for energy are very low. The IT integration in the production creates high costs. New technologies such as hybrid reality or advanced human machine interfaces might improve. The increasing intelligence in components like sensors supports operators in decision making or reduces the effort for people while components decide autonomously. Cyber-physical production systems could benefit from distributed intelligent components in regard to speed and reliability. The complexity of systems is complicating their usage and new standards might help to manage this situation. For instance, the Internet of Things could standardize communication of cyber-physical production systems. These research trends lead to huge numbers of publications.

With regard to this research some aspects of cyber-physical systems are of particular relevance: The quality of production can be measured efficiently by existing performance indicators independent of cyber technologies as shop-floor performance matters. However, there might be different performance indicators illustrating the ability to decide correctly at an opportune time like the autonomy index [27]. Metrics on architecture of manufacturing systems based on the number of sensors versus the ability of reconfiguration of automated systems are presented in [46].

Research on changeable production systems
The first cluster represents all work which emphasize the direction of manufacturing planning. In this context, the question of changeability and adaption of production concepts play a key role. To avoid large investments in automated systems following a product changeover on the shop-floor, the manufacturing systems must be adaptable.

Mecatronic modules are the basis of a methodology presented by [49] which can be used for modelling and evaluation of concepts for manufacturing systems. This methodology assists the engineers in reusing modules. This ultimately reduces engineering effort. [30] use the simulation approach to evaluate the evolution of a industrial plant. Different modification strategies on basis on typical evolution categories were depicted and engineers can synchronize their information with other disciplines.

Metrics like performance indicators are used very often in literature to measure criteria e.g. the project success [4]. Cost, time and quality are basic indicators in nearly every field. Performance indicators measure how well an organization is achieving a particular objective, e.g. a production activity. In industrial manufacturing, performance indicators such as quality, availability, efficiency and numerous other indicators are well established. A literature review about performance measurement is given by [33].

There are a large number of standards and guidelines defining performance indicators for production, for example the ISO 22400 standard. Performance indicators measure how well an organization is achieving a particular objective, e.g. a production activity.

Wiendahl et al. defined changeability as "characteristics to accomplish early and foresighted adjustments of the factory's structures and processes on all levels to change impulses economically" [52]. These studies in the production management domain have identified various meanings of changeability inside a factory. This leads to various definitions such as agility, transformability, flexibility, re-configurability which can be associated with different levels of change within production.

Multiple concepts were developed in the last decades to increase the changeability mostly in the production level. The academic view on flexibility and changeability of production systems was defined by [11]. They distinguish universality, scalability, modularity, mobility and compatibility as enablers for factory changeability.

Product customisation, production with nearly lot size one as well as fast developing manufacturing and assembly technology will lead to a shorter lifecycle of automated systems. Unless these are able to handle the production of new and previously unscheduled products and production processes [5, 31]. Subsequently [34] has raised the question whether this can be evaluated and which methods are suitable.

Research on flexibility due to autonomy of automation equipment
Gronau et al. [14] identifies autonomous cooperation and control as key factor to ensure changeability. Systems should be enabled to react to changes within

boundary conditions and adjust themselves in an appropriate way. For this purpose, important aspects are decision-making, autonomy, interaction and hierarchy.

Further, changeability is seen dependent on indicators such as scalability, modularity, availability, independence, interoperability, self-organization, and self-similarity [53]. Windt [53] highlights the limits of self-control in terms of responsiveness or runtime decision-making in dynamic systems. A multi-component evaluation system was created to define the degree of self-control, depending on complexity of the considered logistics system and the logistical achievement. [41] investigates the potential of self-control as enablers for planning and implementation of changeable production systems.

Weyrich et al. [50] discusses the demand for more flexibility and presents a classification which can be used to assess the flexibility of existing machinery.

Ruiu et al. developed the Potsdam Change Capability Indication which is a knowledge management tool based on use of creativity techniques. Change capabilities are therefore seen as a strategic success factor. In order to break existing thought schemata and conceive new ideas, strategy cards are designed to foster creativity of a group to find a solution and empower e.g. a production system [39]. Zuehlke [55] reports that wireless communication systems reduces the cabling effort and enable flexibility in the layout of plants.

4 Approach and Methodology

Would automation systems provide multiple systems characteristics or features to enable the value-adds mentioned in the introduction?

Literature indicates various abilities providing the proposed value-adds. But, which abilities are necessary to enable the value-adds? And, how are these abilities quantified?

Firstly, visionary and forward looking literature and conference contributions on cyber-physical production systems have been extensively analysed, discussed and brainstormed among the experts. As a result, technology trends in the sense of enabling technologies were conceived. This results in a set of so-called "abilities". The "abilities" are headings for fundamental technologies which enable cyber-physical production systems. Then each ability was categorized to enable a measurement. The measurement will represent degree of realization of these abilities in production systems.

These abilities are potentially shaping cyber-physical production systems and are making them different from ordinary, conventional production systems once available. Abilities are characteristics or features of a system which can be categorized and described in an abstract way. For the evaluation of production systems, performance indicators are a well-established approach.

Therefore and secondly, international standards and literature on production systems and manufacturing management have been reviewed and are full of performance indicators and metrics out of which a subset has been identified. These performance indicators and corresponding metrics have been summed up.

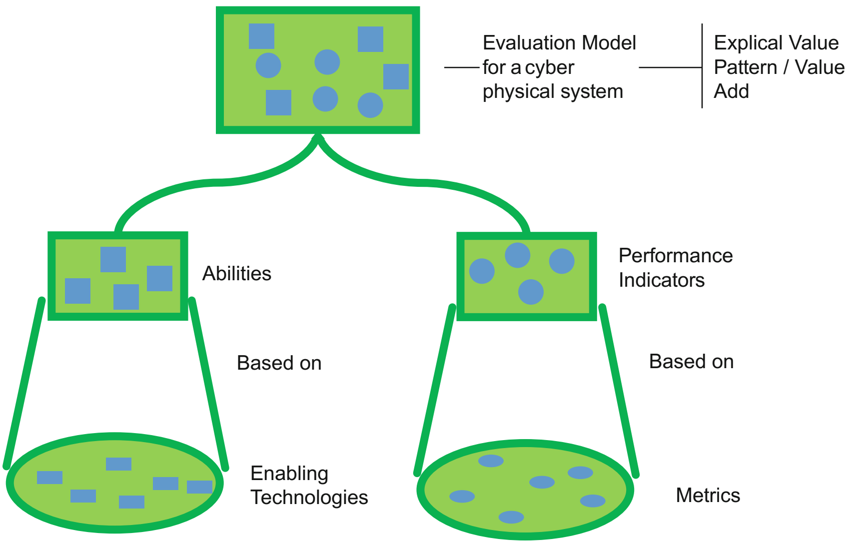


Fig. 4 Methodology of the evaluation model for cyber-physical systems

Thirdly, workshops with experts have been conducted in order to interconnect the conceived potential benefits, performance indicators and metrics. In this creative exercise the experts used brainstorming techniques and a scenario method to define the evaluation model. Hypotheses on potential benefits and the technology drivers were set-up. Experts validated them in discussions using practical examples. During discussion a correlation between abilities and performance indicators for CPPS became apparent. It was found necessary to grade the abilities and correlate them with performance indicators and metrics in order to measure them.

In the final forth step, a set of value patterns for the system characteristics has been defined which aim to sketch the relationship between the "abilities" based on enabling technologies and the "performance indicators" and related metrics.

In order to evaluate the conceived evaluation model, all resulting abilities and performance indicators were clustered using a design structure matrix (DSM) approach. Inconsistencies have been identified and as a result the identification of typical patterns of cyber-physical systems have been derived. The value pattern analysis based on the design structure metrics was useful to structure the argumentation and identify inconsistencies in the model. These inconsistencies were rectified by repeating some discussion of step three.

Figure 4 illustrates the chosen approach. The evaluation model of a cyber-physical system is based on the concept of ability and performance indicators. Abilities are abstractions of technology trends whereas performance indicators are headings for metrics. Based on the evaluation model, value-adds or value patters can be explained.

This methodology was inspired by an approach of foresight based on an advanced scenario method. It is well known in macroeconomics analysis for trend analysis in economy and society [17]. This methodology combined literature surveys group discussions of experts and provides a framework for analysis and projection. The method was enhanced and adopted e.g. with the design structure approach to be used in the study.

5 Abilities of Cyber-Physical Systems

What are the key enablers in the sense of driving technologies of cyber-physical systems? The important question is which characteristics or features should cyber-physical systems possess, and what technology trends are related to them?

The characterization of cyber-physical production systems leads to definition of abilities which are enabling technologies. An automation system can be characterized based on certain technology trends which form potential technical abilities. Ability can be defined as "an acquired or natural capacity or talent that enables [...] to perform a particular job or task successfully" [6]. This capacity enables a technical system to perform particular tasks successfully.

But how are cyber abilities identified? Which research trends or technical developments can be seen as enablers of an ability? Brainstorming among experts was conducted looking at various trend technologies. These trends were accumulated and structured, leading to ability descriptions. For instance cyber-physical production systems abilities would be in the area of capturing and processing of data and information or high quality analysis with artificial intelligence leading to generalization and specialization.

Figure 5 provides an overview of derived abilities and their categorizations. Details of each ability are described by a derivation diagram. The abilities were further grouped and different areas can be distinguished:

- Abilities dedicated to capture and processing of data and information. This relates to a set of abilities which are: data processing ability, networking ability, and IT-Integration ability and perception ability.
- Abilities of high quality analysis with artificial intelligence involve the ability of knowledge creation and reasoning as well as automatic scheduling.
- The specialization ability and generalization ability are relevant for evaluation. Due to the development progress of artificial intelligence these two abilities could not be realized optimally.

However, there are further aspects of cyber-physical production systems as these abilities are not all encompassing. Aside from the mentioned abilities of a cyber-physical production system, other aspects are also of interest. These are distributed architecture, Internet of Things, cloud computing, virtual realization, 3D modelling and physical simulation, seamless integration, new standards, laws and guidelines.

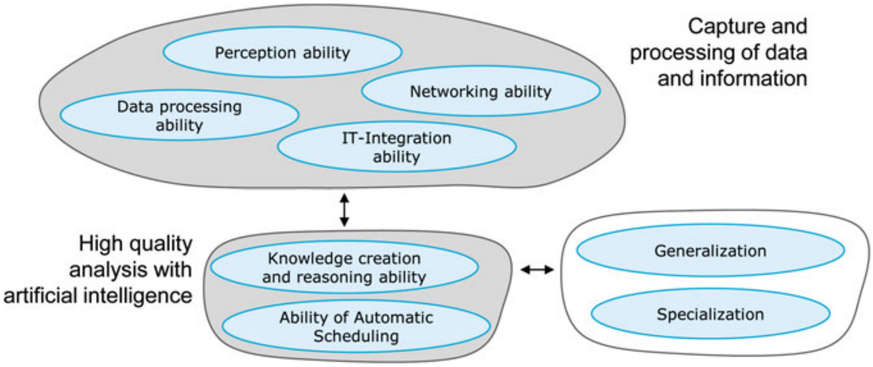


Fig. 5 Grouping of abilities and further aspects

5.1 Abilities of Capturing and Processing Data and Information

The capturing and processing of data and information concludes the ability of a system to handle the relationship between the system itself and its environment. The categorization includes the ability to sense the environment, the ability to process captured data, as well as the ability to network and integrate with other systems. Tables 3, 4, 5 and 6 describe these abilities with possible categorizations, descriptions and examples.

Table 3 Characterisation of data processing ability

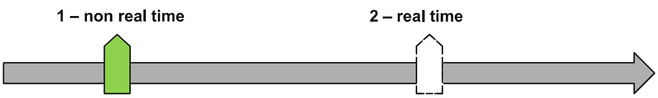
Data processing ability	
Description	An automation system in the future may involve enormous quantities of data, which are complex or change rapidly. The cyber-physical production system should be able to access different databases and process information appropriately. Therefore the *system is required to process acquired data in time*. It is a prerequisite that data is accessible, especially if it is distributed
Categorization	
Example	An example of a five-axis machining station carrying a task of curve interpolation will be taken to depict the different data processing ability
	Category 1 Curve interpolator (including trajectory planning, inverse kinematics transformation, tool position and orientation) is executed with time delay
	Category 2 Curve interpolator is executed without any delay

Table 4 Characterisation of perception ability

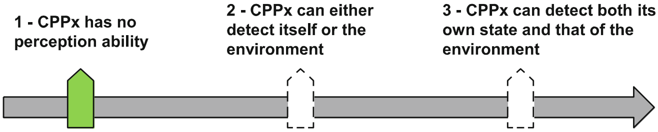
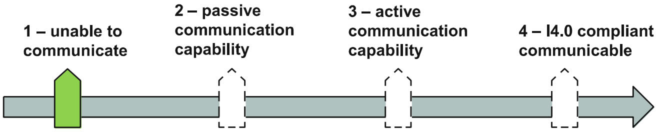
Perception ability	
Description	The perception ability describes *the special perception level of a system of itself or its environment*. It is dependent on the data processing ability and will also be influenced by sensor fusion and the variety of available sensors (e.g. different sensor types with varying sampling frequencies) in global markets (further reading in [26] and [21])
Categorization	
Example	An uneven surface of a work piece was caused by the broken milling tool
	Category 1 The uneven surface is not able to be recognized by the milling work station
	Category 2 The uneven surface is recognized by the milling work station
	Category 3 After recognizing the uneven surface, the work station checks itself and determines that the milling tool was broken

Table 5 Characterisation of networking ability

Networking ability	
Description	The networking ability of a cyber-physical production system describes the *ability of a system to transfer information between different systems*. This ability is derived according to [44], with regards to the communication ability. This ability is influenced by routing technologies, wireless products and standard fieldbuses. (Further reading in (ISO 4930))
Categorization	
Example	An example of a machining station is taken to depict the networking ability
	Category 1 No data exchange between work pieces nor work stations
	Category 2 Work station can read the barcode on a work piece
	Category 3 RFID on a work piece can send signal to work station
	Category 4 Work pieces from different suppliers can exchange data with work stations in different factories

5.2 *Abilities of High-Quality Analysis with Artificial Intelligence*

In addition to the aforementioned abilities, a module or system can be supplemented by artificial intelligence algorithms in order to act more sophisticatedly. Artificial intelligence could create abilities of a system which allows the system to act appropriately in an uncertain environment without the existence of a pre-defined action (Tables 7 and 8).

Table 6 Characterisation of IT-integration ability

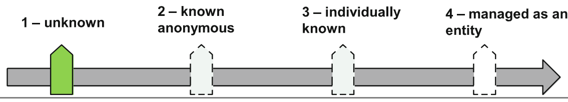
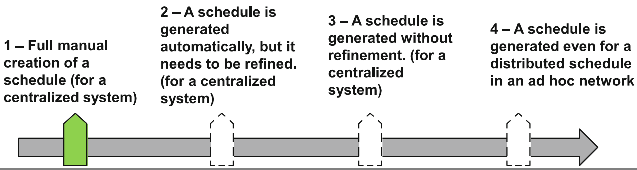
IT integration ability	
Description	Any IT system has to be *integrated into the overall system architecture*. There are different levels to which extend such an integration is implemented. This ability is derived according to [44], too
Categorization	
Example	A drilling work station is connected with a milling work station
	Category 1 The drilling work station is not able to recognize the milling station
	Category 2 The drilling work station detects a milling work station without acquiring its name (series) and functionality
	Category 3 The drilling work station can access the functionality of the milling work station
	Category 4 The drilling work station can managed the drilling work station as an entity

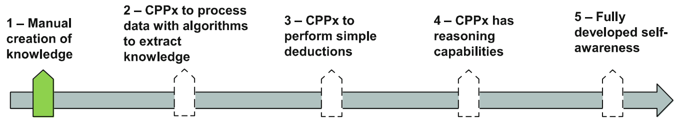
Table 7 Characterisation of ability of automatic scheduling

Ability of automatic scheduling	
Description	Scheduling is the ability to prepare an executable manufacturing plan by assigning resources for manufacturing. The origin of this ability is justified in the manufacturing process management and the available practical scheduling algorithms
Categorization	
Example	A machining work station is required to process a new type of work pieces
	Category 1 A new schedule needs to be generated manually
	Category 2 According to the BOM, a new schedule can be generated digitally by the work station, but it has to be refined manually
	Category 3 According to the BOM, a new schedule BOP can be generated without further refinement
	Category 4 A new schedule for overall plant and multiple BOMs is able to be generated to support BOPs

6 Advanced Abilities

Advanced abilities are divided into the "Industrie 4.0" component and Artificial Intelligence which will be described in the following paragraphs.

Table 8 Characterisation of knowledge creation and reasoning ability

Knowledge creation and reasoning ability	
Description	Knowledge creation and reasoning ability describe the *ability of an intelligent system to create its knowledge to understand its environment or to access its knowledge* and thereby even understanding the reason for a problem and find a solution
Categorization	
Example	An example of examining the quality of work pieces is taken to depict the knowledge creation and reasoning ability
	Category 1 User defines rules for classifying work pieces into different categories
	Category 2 A work station learns some samples, so that it can classify work pieces
	Category 3 For a new type of work pieces, which haven't been learned, a work station can deduce a proper categorization
	Category 4 According to the appearance of a work pieces, the work station can create reasoning
	Category 5 A work station has self-awareness similar to humans

6.1 The "Industrie 4.0" Component

An idea of so-called "Industrie 4.0" component was conceived by a group of experts [44] which published a classification for cyber-physical systems categories. The idea is depicted in the following figure.

According to this VDI initial classification, such a component should possess the communication ability. Furthermore, the IT system should be IT integrated, in the sense of being individually known or managed as an entity.

Figure 6 illustrates the "Industrie 4.0" component. According to this classification such a component would have to be an entity in the IT-system or would have to be individually known in terms of IT. The communication should be based on an "Industrie 4.0" standard.

The potential impact of an "Industrie 4.0" component and interpretation of the term might lead to an advanced value creation e.g. a new standard for communication and IT integration as the discussion in the community suggests.

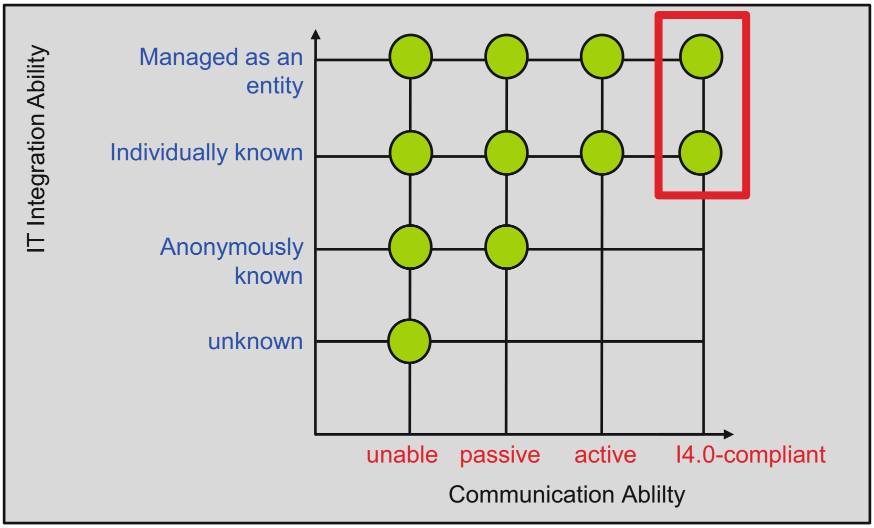


Fig. 6 Classification of an I4.0 component According to [44]

6.2 Artificial Intelligence

More system characteristics which do not exist yet could be outlined utilizing artificial intelligence technologies. For instance, there could be a *specialization ability* which would describe the ability of a system to utilize its knowledge in a concrete situation. A second example would be a *generalization ability* which describes the ability of a system to perform the taxonomic relationship between a general concept and specific instances.

A major step in the transformation of conventional systems into cyber-physical production systems is the integration of Artificial Intelligence functionalities into the modules and systems. Smart modules would have the knowledge about themselves, their abilities and their role inside a cyber-physical production systems production site. This is one of the most important prerequisites to enable intelligent and autonomous cyber-physical production systems. Only these systems would be able to create a decision space wherein they could optimize themselves and the overall system.

In this concept two more abilities are relevant which directly relate to artificial intelligence: The ability to make decisions and act autonomously as well as the ability of self-x. The x can stand for optimization, configuration etc.

Although self-x and autonomy are not identical concepts, they have always been used interchangeably. Self-X is considered to be a life-like property to enable a system to adapt itself to its dynamical environments [3], while autonomy emphasizes the independency of a system to other systems or its user [7]. These two concepts are concerned with the following identical and essential aspects:

- Absence (or extent) of external intervention, controlling or influence to a system
- Changeability (or extent of changeability) of system structure according to the changed goal and/or environment.

Due to this, lots of literature [10, 19, 36] mixes these two concepts. However, autonomy of a system is always associated with Self-X properties, which are important to the implementation of technologies. Based on this research, each individual Self-X (e.g. self-configuration, self-optimization) will be considered to be an aspect of autonomy. The aspects of autonomy can be identified by the following main Self-X properties: self-configuration, self-optimization and self-explaining.

Autonomy is a global property of a system; many factors can affect the degree of autonomy. However in [3, 7, 35], only the decision making is considered to be a factor on autonomy. Obviously knowledge creation and reasoning ability can affect decision making. If a system possesses enough knowledge, it does not require human intervention to propose decisions. Contrarily, if a system does not possess sufficient knowledge, it needs human intervention for decision making. Figure 7 depicts how these relate to each other. There are two additional steps or abilities required before autonomy is reached. The first step is the Knowledge Creation and Reasoning Ability. The ability of processing knowledge in terms of learning and saving rules is a pre-condition for Decision Making. Decision Making is another precondition to autonomy and requires knowledge creation and reasoning ability. This Decision Making assists humans and provides additional information and guidance for decision making.

Undoubtedly a system with higher autonomy can reduce the workload for the operator. Although a fully autonomous system which requires a human-like intelligence is nowadays still difficult to approach, the researcher keeps striving for higher and higher autonomous system (e.g. semi-autonomous system). The quantification of autonomy can be found in [7]. In the first work, autonomy can be calculated based on the size of the inner and outer configuration space.

Autonomy and complexity are two different properties of a system; however they affect each other reciprocally. The cyber-physical production system tends to

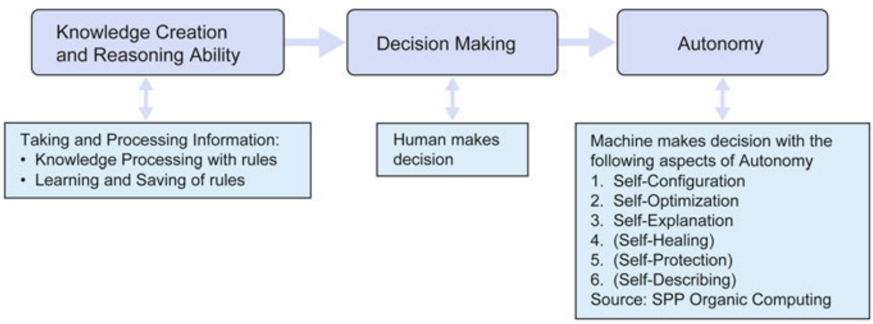


Fig. 7 Knowledge creation, reasoning, decision making and autonomy

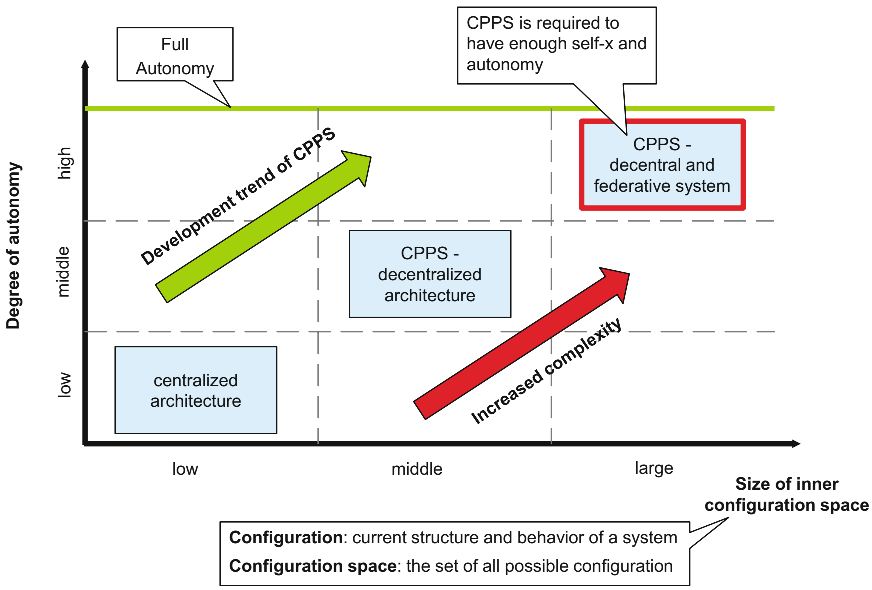


Fig. 8 Autonomy masters the complexity of CPPS

be more and more complex. For example, instead of a decentralized architecture, a cyber-physical production system can be developed as a decentral and federative system comprising of more subsystems and more complex relationships between these subsystems (see Fig. 8).

In this case the size of the inner configuration space enlarges, which meanwhile increases the complexity of a cyber-physical production system. On the other hand, in order to make the subsystems communicate and work well with each other without human intervention, each subsystem is developed to have a certain degree of autonomy. This provides the system the necessary freedom to act without human assistance. In this case humans can neglect the complexity of the system.

7 Performance Indicators

A performance indicator is a type of measure to help compare and support the needs of decision makers.

Definition: Performance indicators measure how well an organization is making progress in achieving a particular objective, e.g. of a production activity. Austin et al. [1, 20, 40, 45] are examples of references on performance indicators.

Success is the achievement of some operational goals in production management.

Numerous performance indicators and other measures are collected and corresponding metrics exist based on multiple definitions in standards, guidelines or best practice in production management.

In manufacturing, performance indicators such as quality, availability, efficiency, instate OEE or downtime are well established in industrial production. There is a large number of standards and guidelines defining performance indicators for production.

A cyber-physical production system is characterized into two levels: the physical and the cyber level. The physical indicators describe the characteristics of physical objects, e.g. hardware of automation systems which is well covered by performance indicators.

The "cyber" aspect might have an impact and enhance the performance indicator.

The quality of a production can certainly be measured efficiently with existing performance indicators independent of any technologies as the shop floor performance matters. However, there might be different performance indicators illustrating the ability to make the right decision at the right time or decide on the architecture of the manufacturing system.

7.1 Identified Performance Indicators

There are several performance indicators influencing the cost of production. Figure 9 depicts the characteristics and associated performance indicators identified for use with cyber-physical production systems.

There are a number of performance indicators identified, which could be grouped into different categories.

The following performance indicators have been identified:

- Modularity, complexity and usability as characteristics of the overall systems architecture
- The usability in the sense of "user friendliness and usability" is associated with the system architecture but includes additional aspects
- The characteristics of changing production system would be indicated by automatic planning as well as a re-configurability indicator.
- Social interaction and support of decision as special "cyber" indicators
- The operation of production is characterized by performance indicators, which are maintainability, production efficiency and automatic adaption.

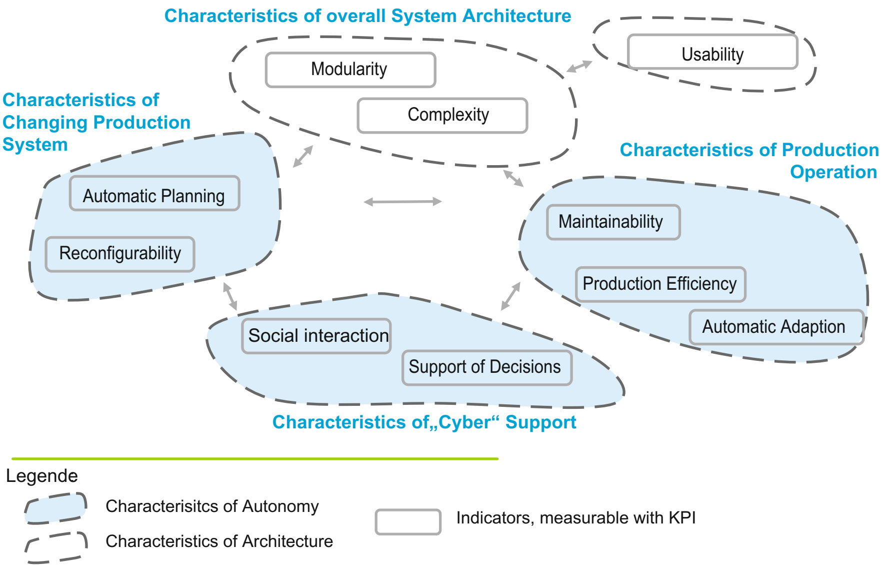


Fig. 9 Classification of measurable indicators towards characteristics of systems

7.2 Performance Indicators of the Overall Systems Architecture

7.2.1 Modularity

Definition: Designing industrial automation systems modularly enables customer specific product variants that can be created in a flexible manner, depending on customer requirements and costs [18]. A module is a unit that is functionally and physically independent from the rest of the system. For industrial automation systems, functions are grouped into individual manufacturing modules.

Impact: Cyber-physical production modules, which are smart modules, are employed in the context of I4.0. The realization of such modules involves more effort than the conventional ones. The value-add of using cyber-physical production modules could be that they allow a more effective and faster diagnosis and reconfiguration, eventually leading to the reduction of rescheduling costs. Modularity affects complexity. It is related to performance indicator for the architecture of manufacturing systems.

Potential Measures are for example the number of modules, the module size, number of required modules, the vector modularity measure [42], Coupling between Objects—CBO, Depth of Inheritance Tree (DIT), Interface Complexity [8]. The optimal quantification of these performance indicators for modularization is difficult to be determined and needs further research.

7.2.2 Complexity

Definition: The term complexity can be used to characterize a system or sub-system with many parts where these parts interact with each other in multiple ways [2, 32].

Complexity appears in multiple forms, e.g. the complexity of the solution in manufacturing, of the design or engineering process. This means there are different complexities of architectures such as hardware or software design, as well as the complexity of tasks during manufacturing.

Impact: Due to a decrease in complexity of tasks of operation or manufacturing, it is possible to reduce costs of rescheduling and decrease the effects on other products for the following reasons:

Cyber-technology such as the ability of automatic scheduling reduces the complexity of tasks and makes reconfiguration, rescheduling and maintenance easier and faster. Thus, the time to convert to the alternative operation is reduced and the rescheduling costs are reduced. Reference architecture and standards [9] could simplify the complexity of the production of different products during the operation as maintenance, configuration or optimization might become easier. However, the process of designing cyber-physical production systems might become more complex as the complexity of the overall system increases compared to conventional systems.

Complexity affects rescheduling cost. It is a performance indicator for the architecture of manufacturing systems.

Halstead Complexity [15] and the assortment complexity are potential metrics used in software development.

7.3 Performance Indicators of Production Operation

7.3.1 Maintainability

Definition: Ease with which a manufacturing system can be maintained.

Impact: Cyber-technology reduces the diagnosis and maintenance times and increases the fault finding rate based on cyber abilities such as data processing ability, knowledge construction ability or reasoning ability. Based on advanced processing features, a fault can be located faster. Also the time for reconfiguration can be reduced. Maintainability affects rescheduling cost and makes a difference in production cost.

Potential Measures are 5 S Status [43], Availability of documentation, Time between Failure and Reconfiguration, Error finding rate, Mean Time To Failure [20], Mean Time To Repair [20].

7.3.2 Production Efficiency

Definition: Productive efficiency is the ability to "succeed in producing an as large as possible output from a given set of inputs" [12].

An industrial system can be measured upon the ratio between the gains and the expenditures. Efficiency can also be viewed as an operational state whereby a company cannot increase output of a specific good or service without additional costs [6].

Impact: In comparison to non-effective production systems, the difference of production costs and the effects on other products are reduced. Production efficiency makes a difference in production cost.

Potential Measures are the Cycle Time, Scrap ratio (ISO 22400), Rework ratio (ISO 22400), First Pass Yield (ISO 22400), Operational Capacity, numbers of delayed products.

7.4 Performance Indicators of Changing Production Systems

7.4.1 Re-configurability

Definition: Capability of a system which allows for changing behaviour of the system by reconfiguration. The re-configurability derives from the systems configurability, i.e. the configuration corresponds to the design, the selection and composition of modules [16].

Impact: Cyber technology reduces the effort for reconfiguration of the production system. It facilitates the execution of the alternative operation. Self-adaptation of a system would help reduce reconfiguration costs, as smaller changes would be done automatically.

Potential Measures are number of reconfigured sub-systems, reconfiguration time, Setup time, and Success or Error rate of an adaptation.

7.4.2 Automatic Planning

Definition: Ability to organize activities to achieve a desired goal.

Impact: Cyber technology enables automatic planning of schedule and any other reconfigurations. This indicator is a far reaching ability of self-organization in individual modules or whole networks. An automatic planning ability affects rescheduling costs and impacts production costs. A prerequisite is a digital model of the production system in order to have relevant information available for automatic planning.

The performance indicator Every Part Every Interval [38] or the planning times are *possible measures*. Additionally the metrics of production efficiency can be used to evaluate the planning result.

7.4.3 Automatic Adaptation

Definition: Ability to adapt activities to optimize a process.

Impact: There are different kinds of adaptation: The system can adapt in the meaning of Self-Optimization and optimize the production efficiency. Furthermore the system can adapt to new requirements in the meaning of Self-Configuration.

Potential Measure is the number of adapted parameters.

7.5 Performance Indicators of Cyber Support

This paragraph shows an overview of the indicators particularly related to the support of decisions. Besides the classical production indicators, these characteristics are in addition to physical performance indicators highlighting the "cyber" aspects.

7.5.1 Social Interaction

Definition: Advanced capability of systems, which can interact with other systems or human (i.e. system user), so that the systems are interested in one another and are able to empathise with other technical systems or users. Cyber-physical production systems could encompass artificial intelligence.

Impact: Cyber technology supports exchange of information, negotiation with another systems or system users, acting in a given environment context. Social interaction facilitates the cooperation between cyber systems. It affects the decision making, Self-Configuration and Self-Optimization.

Potential Measures is the availability of advance human communication ability such as language, semantic definitions, symbol recognition or learning capabilities.

7.5.2 Support of Decisions

Definition: Ability to provide knowledge service or proposals which enable the user to make an informed decision or follow a predefined action.

Impact: Cyber technology enables a new generation of services, providing the production manager with information services and background knowledge. These services can support decision-making. An ability to support decisions affects rescheduling costs, makes a difference in production costs and helps assess the impact on other products. A prerequisite is a digital twin in order to have all required information available as well as elaborated information and knowledge processing methods.

Potential Measures are number of decision-making situations supported, quote of successfully supported situations or similar ratios.

7.5.3 Further Characteristics: Usability

Definition: Usability is the ease of use of a manufacturing system or its sub-systems (see ISO 9241-11)

Impact: Cyber and any related service to support diagnosis and operation management will be beneficial It indicates whether the user can use and operate the system easily. The usability describes the extent and ease to which an operator can deploy the system or service in a given context.

Potential Measures are the success rate (of correct operation), Mean-time a task requires, rate of operating errors and the number of clicks for functions to be executed.

7.6 Relation Between Performance Indicators and Abilities

There is a correlation between performance indicators and abilities. Both were conceived by applying a different methodology but describing the same type of a cyber-physical production system.

Table 9 shows the relation between the performance indicators and the related abilities. The most evident relations were added to the table, however further relations between the performance indicators and abilities can be argued with view to specific use cases.

8 Validation

The analysis so far has delivered a set of cyber-physical production system evaluation model parameters which are the abilities and the performance indicators. They stand for enabling technologies improving a production system or for defined metrics identifying the performance of a system. That means a characterization of the effects based on abilities can be measured by the derived performance indicators. However, are there relevant clusters and fundamentally new approaches of cyber-physical production systems? To understand the relationship between the abilities and performance indicators in more depth both were subjected to an analysis of the correlation yielding clustered patterns.

Table 9 Performance indicators and related cyber abilities

Performance indicators	Related abilities
Modularity	• Communication ability
	• IT integration ability
Complexity	• Ability of automatic scheduling
Usability	• Knowledge creation and reasoning ability
Maintainability	• Knowledge creation and reasoning ability
Production efficiency	• Communication ability
Automatic planning	• Ability of automatic scheduling
Re-configurability	• Ability of automatic scheduling
Automatic adaption	• Knowledge creation and reasoning ability
	• Data processing ability
Social interaction	• Perception ability
	• Communication ability
	• Knowledge creation and reasoning ability
Support of decisions	• Knowledge creation and reasoning ability
	• Perception ability

A clustering analysis is presented based on the approach of Design Structure Matrix (DSM; see: www.dsmweb.org/ more information). This describes the interconnection between performance indicators and abilities. The analysis is based on the defined relationships as per Fig. 5: Grouping of abilities and further aspects of a cyber-physical production systems. The resulting DSM is displayed in Fig. 10. The matrix was optimized and four clusters were identified:

It is evident from the DSM that there are many cross correlations between the abilities and performance characteristics which cannot fully be assigned to a certain cluster. However, four patterns can be identified clearly.

Pattern 1—"Smart Modules":

The performance indicator of Modularity is clearly driven by two technical abilities; the Communication Ability and the IT Integration Ability. This relationship gravitates around an "Industrie 4.0" Component.

Pattern 2—"Self-configuration":

The performance indicators Automatic Adaptation, Automatic Planning and Re-configurability are mainly driven by the ability of automatic scheduling. The performance indicators of re-configurability and automated adaption are assigned to the characteristics of self-configuration. These indicators can show how easily a production system can be modified to a new configuration of hard- and software.

Pattern 3—"Self-optimization":

The performance indicator of Product Efficiency is linked to Perception Ability and Data Processing Ability. If a system is able to perceive its own status and process related data, it can improve its own performance leading to Self-Optimization. Another aspect in the context of *self-optimization* characteristics includes the

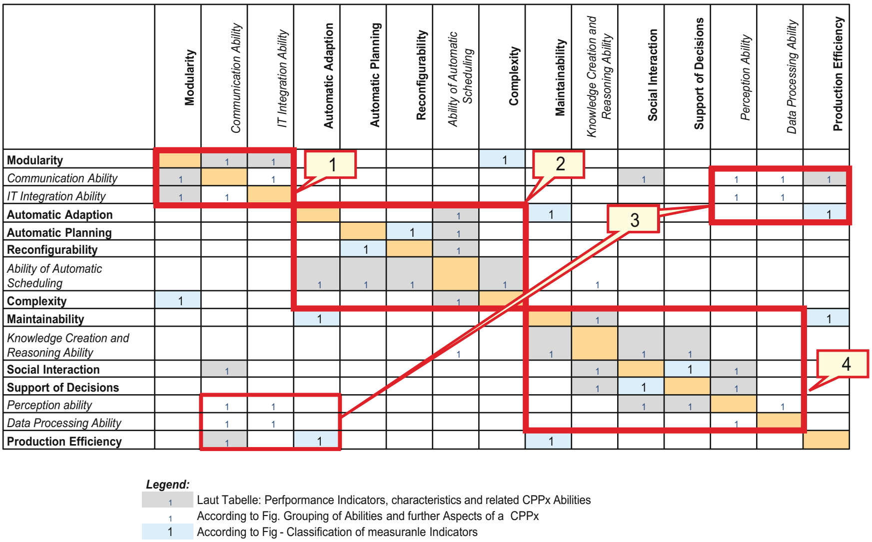


Fig. 10 Clustered DSM matrix of measurable performance indicators and technological abilities (the following Abilities are not displayed for the reason of simplicity: generalization ability, specialization ability and usability)

production efficiency, maintainability and automatic planning. These can be used to measure the operational performance of a manufacturing system.

Pattern 4—"Decision Support":

The performance indicator of Decisions, Social Interaction and Maintainability are driven by the Ability of Knowledge Creation and Reasoning as well as a Perception and Data Processing ability. This is evident as the processing of data and information in a networked system would consequently support any decision making. The performance indicators deployed are the support of decisions and social interaction related to the provision of information and knowledge.

Further validation using simulation studies of cyber-physical production systems are planned. So far the logical consistence of the derivation of abilities and performance indicators towards typical value patterns could be proven.

9 Conclusion

The motivation to utilize cyber-physical systems is supported by a value-add analysis which systematically collected statements from visionary publications. The resulting list of potential benefits of cyber-physical systems has been condensed from the perspective of automation and information technology. After a thorough review of the literature on evaluation methods for production systems, a qualitative evaluation model was conceived describing "cyber" abilities and performance indicators.

A set of characteristic has been defined based on literature research and brainstorming, explaining the impact in terms of enabling technological abilities and performance indicators for cyber-physical production systems.

The defining technological abilities can be grouped in two areas:

- Capture and processing of data and Information, which include Data Processing, Communication, Perception and IT Integration Ability
- The Artificial Intelligence, supported by abilities for the Knowledge creation and reasoning, Automatic Schedule and Generalization and Specialization

These abilities aim to provide a descriptive model explaining cyber-physical production systems from a viewpoint of technology. Different levels of systems empowerment could be defined but the availability of these technologies have no directly measurable correlation to the improvement of production performance.

Further, with regard to measurable system characteristics, an intensive study of standards and guidelines has been conducted. A large set of conventional metrics on production and manufacturing management is available and was extended to include the newly developed performance indicators as well. This part of the evaluation model is based on a set of performance indicators which can be measured directly or indirectly by means of metrics in the production. The performance indicators which are important for cyber-physical production systems can be categorized in two classes:

- Existing and known performance indicators, like Modularity, Complexity, Usability or Maintainability
- There are also new and important system characteristics which are Automatic Adaption, Social Interaction and Support of Decisions.

Related metrics have been identified in order to measure these characteristics. A design structure analysis has been conducted, which displays the correlation between the identified abilities and performance indicators. As a result, value patterns can be identified based on enabling technologies described by the abilities, and their effects described by the performance indicators.

Four patterns revile important relations for the evaluation model:

Modularity, Communication ability and IT Integration Ability form the "Smart Modules" pattern in the sense of a new generation of module with high level of IT integration and standardized communication. The "Self-configuration" pattern is based on automatic scheduling, planning and adaptation, to support autonomous re-reconfiguration. The third pattern "Self-optimization" is based on autonomous optimization technologies to increase Product Efficiency. The last pattern "Decision Support" or self-explanation of cyber-physical production systems is based on data processing, perception, knowledge processing and reasoning, social interaction e.g. to optimize Maintenance. These patterns demonstrate the correlation of the different aspects for the evaluation model.

10 Further Work

A further validation of the developed evaluation system, studies or simulation experiments needs to be carried out. Discussion with experts, e.g. from the discrete manufacturing or process automation domain could support the application of the evaluation model towards different scenarios.

Further work is required to explain the results, e.g. the rather abstract definitions of the technological abilities and performance indicators. It was possible to conduct studies on automotive manufacturing in powertrain. Though the results were very encouraging, further investigations would be required to refine the description models further and breakdown its characteristics to the domain.

Further, scenario experiments e.g. based on simulation are necessary for metrics to measure the value-add of self-configuration and self-optimization. The dynamic interaction of cyber-physical production systems in terms of local and global optimization also needs to be investigated to understand self-optimization functions in more depth.

References

1. Austin RD, DeMarco T, Lister TR (1996) Measuring and managing performance in organizations. Dorset House Publishing, New York
2. Baccarini D (1996) The concept of project complexity–a review. Int J Project Manage 14:201–204
3. Ball M, Callaghan V (2012) Explorations of autonomy: an investigation of adjustable autonomy in intelligent environments. In: Proceedings of the 8th international conference on intelligent environments (IE), Guanajuato, Mexico, 26–29 June 2012, IEEE. doi:10.1109/IE.2012.62
4. Belassi W, Tukel OI (1996) A new framework for determining critical success/failure factors in projects. Int J Project Manage 14(3):141–151. doi:10.1016/0263-7863(95)00064-X
5. Bi ZM, Lang SY, Shen W, Wang L (2008) Reconfigurable manufacturing systems: the state of the art. Int J Prod Res 46(4):967–992
6. BusinessDictionary.com: Production Efficiency. WebFinance, Inc. http://www.businessdictionary.com/definition/production-efficiency.html. Accessed 09 Apr 2016
7. Cakar E, Mnif M, Müller-Schloer C, Richter U, Schmeck H (2007) Towards a quantitative notion of self-organisation. In: Proceedings of the IEEE congress on evolutionary computation (CEC 2007), Singapore, Singapore, 28–28 September 2007. doi:10.1109/CEC.2007.4425022
8. Chidamber S, Kemerer C (1994) A metrics suite for object oriented design. IEEE Trans Softw Eng 20(6):476–493. doi:10.1109/32.295895
9. DKE VDE (2015) German standardisation roadmap Industrie 4.0. DKE, VDE: Die Deutsche Normungs-Roadmap Industrie 4.0. Version 2.0 (Stand October 2015)
10. Dryden J (2015) Autonomy. In: Internet encyclopaedia of philosophy—a peer reviewed academic resource. http://www.iep.utm.edu/autonomy/ Accessed 25 Oct 2014
11. ElMaraghy HA, Wiendahl HP (2009) Changeability—an introduction. In: ElMaraghy HA (ed) Changeable and reconfigurable manufacturing systems. Springer series in advanced manufacturing. Springer, London, pp 17–18
12. Farrell MJ (1957) The measurement of productive efficiency. J Roy Stat Soc Ser A (General) 120(3):253–290
13. Geisberger E, Broy M (eds.) (2012) Agenda CPS: Integrierte Forschungsagenda Cyber-Physical Systems, vol 1. Springer, Heidelberg

14. Gronau N, Lämmer A, Andresen K (2007) Entwicklung wandlungsfähiger Auftragsabwicklungssysteme. Wandlungsfähige ERP-Systeme: Entwicklung, Auswahl und Methoden, p 45
15. Halstead MH (1977) Elements of software science. Elsevier, Amsterdam
16. Heisel U, Meitzner M (2006) Progress in reconfigurable manufacturing systems. In: Reconfigurable manufacturing systems and transformable factories. Springer, Berlin, pp 47–62
17. Hofmann S, Rollwagen I, Schneider S (2007) Germany 2020—new challenges of a land on expedition. https://www.dbrsearch.com/PROD/DBR_INTERNET_ENPROD/PROD0000000000021082/Germany_2020_New_challenges_for_a_land_on_expedi.pdf. Accessed 16 Oct 2015
18. Hohnen T, Pollmanns J, Feldhusen J (2013) Cost-effects of product modularity–an approach to describe manufacturing costs as a function of modularity. In: Smart product engineering. Springer, Berlin, pp 745–754
19. ISO 13482:2014: Robots and robotic devices—safety requirement for personal care robots. Autonomy: “Ability to perform intended tasks based on current state and sensing, without human intervention”
20. ISO 22400-2:2014: Automation systems and integration—key performance indicators (KPIs) for manufacturing operations management
21. ISO/IEC 15438: (ISO/IEC 15416) Information technology—automatic identification and data capture techniques
22. Jazdi N (2014) Cyber-physical systems in the context of industry 4.0. In: IEEE international conference on automation, quality and testing, robotics, Cluj-Napoca, Romania
23. Kagermann H, Wahlster, W, Helbig J (2013) Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0. Abschlussbericht des Arbeitskreises Industrie, 4
24. Kagermann H, Helbig J, Hellinger A, Wahlster W (2013). Recommendations for Implementing the strategic initiative INDUSTRIE 4.0: securing the future of German manufacturing industry; final report of the Industrie 4.0 working group. Forschungsunion
25. Kennzahlen. www.vdma.org/documents/…/63acfab3-4546-4e1a-b396-1551f53f9c32. Accessed 15 Oct 2015
26. Klocke F, Backmeyer M, Blattner M, Eisenblätter G, Jamal R, Joseph Y., Trächtler A (2014) Sensoren in der digitalen Produktion. In: Brecher C, Klocke F, Schmitt R, Schuh G (eds) Industrie 4.0: Aachener Perspektiven: Aachener Werkzeugmaschinenkolloquium 2014, 1st edn. Shaker, Herzogenrath, pp 271–96
27. Lass S, Theuer H, Gronau N (2012) A new approach for simulation and modelling of autonomous production processes. In: Proceedings of the 45th Hawaii international conference on system science (HICSS), Maui, Hawaii, 4–7 January 2012. IEEE
28. Lee EA (2006) Cyber-physical systems—are computing foundations adequate. Position paper for national science foundation workshop on cyber-physical systems, 16 October 2006, Austin, TX, USA
29. Lee EA (2008). Cyber physical systems: design challenges. In: 2008 11th IEEE international symposium on Object oriented real-time distributed computing (ISORC), May 2008. IEEE, pp 363–369
30. Legat C, Steden F, Feldmann S, Weyrich M, Vogel-Heuser B (2014) Co-evolution and reuse of automation control and simulation software: identification and definition of modification actions and strategies. In: IECON 2014—40th annual conference of the IEEE industrial electronics society, pp 2525–2531
31. Leitao P (2009) Agent-based distributed manufacturing control: a state-of-the-art survey. Eng Appl Artif Intell 22(7):979–991
32. Mitchell M (2009) Complexity a guided tour. Oxford University Press, Oxford, p 366
33. Neely A, Gregory M, Platts K (2005) Performance measurement system design: a literature review and research agenda. Int J Oper Prod Manage 25(12):1228–1263. doi:10.1108/01443570510633639
34. Nyhuis P, Fronia P, Pachow-Frauenhofer J, Wulf S (2009) Wandlungsfähige Produktionssysteme–Ergebnisse der BMBF-Vorstudie “Wandlungsfähige Produktionssysteme”. wt Werkstattstechnik

online 99 (2009) Nr. 4, pp 205–210. Internet: www.werkstattstechnik.de. Springer-VDI-Verlag, Düsseldorf
35. Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybern Part A Syst Hum 30(3):286–297. doi:10.1109/3468.844354
36. Piper M (2015) Autonomy (normative). In: Internet encyclopaedia of philosophy—a peer reviewed academic resource. http://www.iep.utm.edu/aut-norm/. Accessed 25 Oct 2014
37. Porter ME, Heppelmann JE (2014) How smart, connected products are transforming competition. Harv Bus Rev 92(11):11–64
38. Rother M, Harris R (2001) Creating continuous flow. Lean Enterprise Institute, Brookline
39. Ruiu M, Ullrich A, Weber E (2012) Change capability as a strategic success factor-a behavior pattern approach for operationalizing change capability. In: Proceedings of the conference on strategic management, pp 23–31
40. Samsonowa T (2012) Industrial research performance management: Key performance indicators in the ICT industry. Contributions to management science. Physica-Verlag, Heidelberg
41. Scholz-Reiter B, Sowade, S (2010) Der Beitrag der Selbststeuerung zur Wandlungsfähigkeit von Produktionssystemen. http://www.sfb637.uni-bremen.de/pubdb/repository/SFB637-B2-10-005-IIC.pdf, Accessed 16 Oct 2015
42. Stryker AC (2010) Development of measures to assess product modularity and reconfigurability. Dissertation, Wright-Patterson Air Force Base, Ohio
43. Takeda H (2006) Das synchrone Produktionssystem. Just-in-Time für das ganze Unternehmen; 5, aktualisierte Auflage; Verlag Moderne Industrie; Landsberg
44. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik (2014) Statusbericht; Industrie 4.0; Gegenstände, Entitäten, Komponenten. VDI e.V., Düsseldorf. https://www.vdi.de/fileadmin/vdi_de/redakteur_dateien/sk_dateien/VDI_Industrie_4.0_Komponenten_2014.pdf. Accessed 9 Oct 2015
45. VDMA-Einheitsblatt 66412-1 (2009) Manufacturing Execution Systems (MES)
46. Vogel-Heuser B, Weber J, Folmer J (2015) Evaluating reconfiguration abilities of automated production systems in Industrie 4.0 with metrics. In: Proceedings of the IEEE 20th conference on emerging technologies and factory automation (ETFA 2015), Luxembourg, Luxembourg, 8–11 September 2015. IEEE
47. Wahlster W (2014) Industrie 4.0-Künstliche Intelligenz in der Produktion, Konferenz zu Ideen 2020, 03 February 2014, Saarbrücken
48. Wahlster W (2014) Normung und Standardisierung-Schlüssel zum Erfolg von Industrie 4.0. In: Workshop Plattform I40 und DKE, 18 February 2014, Berlin
49. Weyrich M, Klein P, Steden F (2014) Reuse of modules for mechatronic modeling and evaluation of manufacturing systems in the conceptual design and basic engineering phase. In: IFAC world
50. Weyrich M, Wior I, Bchennati D, Fay A (2014) Flexibilisierung von Automatisierungssystemen: Systematisierung der Flexibilitätsanforderungen von Industrie 4.0; wt-online 3–2014, pp 106–111
51. Weyrich M, Diedrich C, Fay A, Wollschläger M, Kowalewski S, Göhner P, Vogel-Heuser B (2014) Industrie 4.0 am Beispiel einer Verbundanlage. Automatisierungstechnische Praxis (atp) 56(7):52–61
52. Wiendahl H, ElMaraghy HA, Nyhuis P, Zäh MF, Wiendahl H, Duffie N, Brieke M (2007) Changeable manufacturing—classification, design and operation. CIRP Ann Manuf Technol 56(2):783–809. doi:10.1016/j.cirp.2007.10.003
53. Windt K (2008) Ermittlung des angemessenen Selbststeuerungsgrades in der Logistik-Grenzen der Selbststeuerung. In: Nyhuis P (ed) Beiträge zu einer Theorie der Logistik. Springer, Berlin, pp 349–372
54. Windt K, Hülsmann M (2007) Changing paradigms in logistics—understanding the shift from conventional control to autonomous cooperation and control. In: Understanding autonomous cooperation and control in logistics. Springer, Heidelberg, pp 1–16
55. Zuehlke D (2010) SmartFactory—towards a factory-of-things. Ann Rev Control 34 (1):129–138. doi:10.1016/j.arcontrol.2010.02.008

Part III
Architectural Design Patterns for CMS and IIoT

CPS-Based Manufacturing with Semantic Object Memories and Service Orchestration for Industrie 4.0 Applications

Jens Haupert, Xenia Klinge and Anselm Blocher

1 Introduction

Driven by the individual customers' increasing demands and needs of mass customization in small lot sizes (down to lot size 1) of products, Industrie 4.0—the so-called fourth industrial revolution—brings a major paradigm change to the entire industrial sector. Technologies developed in the field of the Internet of Things and the Internet of Services [1] provide the technical infrastructure to enable mass customization with near mass production efficiency and reliability [2]. Leading to smart products, smart services, and smart data, mass customization creates a competitive advantage in the industrial economy, the service economy, and the emerging data economy [3]. The digital value added in the means of production, products, and systems will enable continuous improvement in the performance of industrial processes ("vertical integration") in manufacturing, engineering, supply chain, life cycle management as well as in the horizontal integration beyond company boundaries. The first steps in this direction have been done in optimizing production processes on different levels of enterprise control systems, e.g. at the management execution system (MES) level [4]; but this is just the beginning.

The key enablers of this disruptive change in the traditional, centrally controlled production process are decentralized embedded cyber-physical systems (CPS)

J. Haupert (✉) · Xenia Klinge
German Research Center for Artificial Intelligence (DFKI GmbH),
Projektbuero Berlin, Alt-Moabit 91c, 10559 Berlin, Germany
e-mail: jens.haupert@dfki.de

Xenia Klinge
e-mail: xenia.klinge@dfki.de

A. Blocher
German Research Center for Artificial Intelligence (DFKI GmbH),
Campus D3_2, 66123 Saarbruecken, Germany
e-mail: anselm.blocher@dfki.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_8

networked together via internet technologies, that allow products and assets to use machine-to-machine communication (M2M) to mutually exchange information about their status, their environment, the stage of production or maintenance, to initiate actions, and interactively control each other. Equipped with active digital product memories with embedded sensors and actuators, Smart Products become CPS and exert an active influence on the manufacturing processes: The products know which components have already been integrated and which production steps still have to be run through, what transport and storage is required. They transmit the critical signals for the downstream production processes, which is stored in their digital product memories. The product under construction controls its own manufacturing process, monitors over the embedded sensors the relevant environmental parameters, and initiates an appropriate response to a fault signal—it is both an observer and an actuator at the same time. The production always maintains the pace of the human who can step in at any time.

The necessary flexibility, adaptability, agility, and interoperability can only be achieved by the use of semantic technologies. They describe in a generic and machine-understandable representation in ontologies the characteristics, functions, and processes of the manufacturing processes and all actors involved. The aim is to achieve a highly adaptable manufacturing process at all levels of the production system—from planning and control to operation while interacting with processes, products, production resources, and employees.

With the Object Memory Model (OMM) we have designed a generic framework in a W3C incubator group for implementing active semantic product memories [5]. The information for the product memory entries is ideally encoded in Semantic Web languages like RDF or OWL, so that a machine-understandable ontology and standardized epistemological primitives can be used for automatic processing. Semantic technologies embedded into OMM [6] guarantee interoperability of the product memory during the complete lifecycle of smart products and enable ubiquitous access by smart data analytics, smart services, and end users to the smart product's lifelog [3].

Dynamic, decentralized production processes offer an approach to a more efficient use of resources: Events like interruptions in the production cycle, raw materials of varying quality, energy or material bottlenecks are detected and corrected in a timely manner by decentralized sensors. Friction losses from the flow of information through a central control unit are minimized. The resource requirements can be determined and planned dynamically: Resources like water, power, or raw materials can be delivered on demand, thereby reducing excess capacity. Information and Communication Technologies (ICT) become a catalyst for a more conservative use of resources in production.

A CPS-based factory with its distributed control system allows flexible control of the components and dynamic orchestration of the production processes up to "plug'n'produce", the rapid replacement with forthwith operational readiness of individual components of assets during the production process. Dynamic orchestration describes the generation of precise, executable processes immediately before their execution by semantically comparing the requirements of an abstract process description and the (smart) services available in a production unit. As core component of this dynamic orchestration a semantic service architecture based on a production ontology and ubiquitous M2M-communication realizes intelligent semantic matchmaking between emerging products and production tools [7], i.e., discover suitable (smart) services, evaluate them and finally select the most appropriate service in the given context.

The adaptable, reconfigurable production (and production unit) for the optimized manufacture of variant-rich, customized products in the context of high-mix, low-volume manufacturing, stands in a hybrid arrangement of manufacturing equipment and human labor supported by new, customized, production assistance systems. Workers, plant operators, and team leaders can be assisted in the execution of specific tasks like assembly planning, integrated dynamic detailed planning of the production processes, employee scheduling, fault management, quality assurance and workplace adaptations by specific assistance systems and IT tools.

In the next section a use case scenario is presented as a frame for the implementation of Digital Object Memories within an industrial setting. Section 3 then examines related concepts and technologies to put the ideas presented in Sects. 4 and 5 into context. While Sect. 4 focuses on the object memory model and its inner workings, Sect. 5 zooms out to a wider perspective and proposes a solution for a smart factory production line as pictured in the use cases. Section 6 completes this chapter with a conclusion and an outlook.

2 Use Cases

A chain of connected use cases serves as a concomitant example to help putting the presented technologies in context. This first description is therefore confined to a conceptual level, introducing the processes and participating entities themselves, while a more detailed description of the technologies featured in the scenario is given over the course of the chapter with their respective introductions.

The use cases describe a metal process workshop in which workpieces are manufactured according to instructions provided by customers. A central application for Service Orchestration overviews the whole workshop in which all relevant physical entities, such as machines and workpieces are represented digitally by Semantic Object Memories.

This scenario is prototypical for similar production lines, as many of them feature tools and materials which share common definitions and only vary in specific parameters. As the concrete specification of these objects relies predominantly on their production context, no general proposition can be made about the design of their digital representation. Therefore, the examples developed in the use cases are restricted to this specific scenario. The demonstrated possibilities and modeling approaches, however, can prolifically be transferred to other production contexts. In a cyber-physical production environment enriched with Semantic Object Memories and Service Orchestration, arbitrary workpieces can be processed automatically in variably sized batches down to individual orders without delay, additional expenses or any other adjustment. The production itself, as well as adjacent problems, such as quality management or maintenance planning can be addressed by the technologies presented in the following sections.

First, a short description of the use case layout is given which then leads to the presentation of three possible scenarios within this layout.

The metal production line has four different stages, connected by the central Service Orchestration application which is able to interact with the digital representations of the physical objects. Upon arrival the workpieces are scanned for their properties and instructions and then processed accordingly. The workshop has a drilling floor with a collection of different drilling machines, as well as a milling floor with a correspondent collection of milling cutters, both of which are optional stages in a workpiece's processing (as not every piece requires both, though at least one). Lastly, the finished items are passed on to a new destination outside of the workshop.

The core piece of this scenario is the Service Orchestration unit overseeing and guiding the whole process. In order to automatically devise production, it needs to have knowledge about the workshop, its capabilities, properties and contents, but also about the workpieces and their current and intended states. This knowledge is provided by Semantic Object Memories, digital representations of the devices and materials within the production line.

Object Memories

For each arriving workpiece an object memory is retrieved, specifying its qualities for the orchestration unit. The memory contains information about its physical state, such as size, material and shape, as well as processing instructions, quality check results and current production history. The composition of such a memory is exemplified in Fig. 1. In detail, the data needed in the use cases comprises:

- **Production instructions**. These are a collection of operations to perform on the object, parameters to these operations and an optional sequential order.
- **Priority**. This entry might be used to steer a workpiece's path through the production line, following specific requirements such as fast production or energy efficiency.
- **Shape**. The object's dimensions are relevant for choosing a correctly sized machine, as well as interpreting the given working instructions.

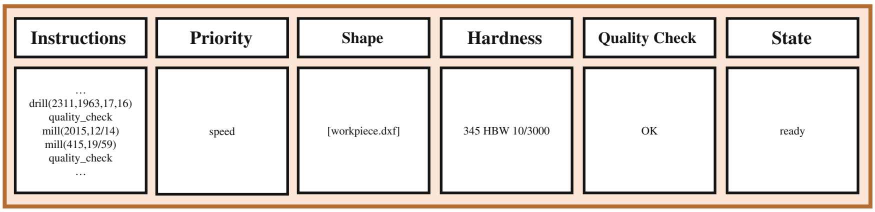


Fig. 1 Example for a possible workpiece's memory contents

- **Hardness**. Some tools are restricted in their capacities concerning especially hard materials and can be damaged if forced to work on harder objects.
- **Quality check**. Diagnostic information about the workpiece's progress after one or more instructions have been executed.
- **State**. This parameter is used by the Service Orchestration to determine the next production step in conjunction with the instructions.

Each individual object of course actualizes the information specifically, but all follow this model and utilize a predefined set of ontology-based values, so that the Service Orchestration is able to revert to a common semantic when processing the data. For example, an object's hardness can be given in different measurement units, such as Rockwell or Brinell, so the used unit has to be either fixed in advance or given as an additional information to the hardness value. The shape could either be a vector of lengths for simpler objects or a full-fledged 3D model for more complex ones. The instructions might be realized as a sequence of functions with parameters, such as "drill(2311, 1963, 17, 16)", where the functions and their parameters are part of the ontology.

The digital representations of the different available machines follow a common memory model as well and fill it with their specific information in agreed-on formats. Figure 2 shows an example for such a memory. Here, the necessary data is complementary to that of the workpieces in some degree:

- **Function**. The (ontology-based) description of the function this device is able to perform.
- **Properties**. Minimum and maximum sizes, shape and degrees of hardness the device is able to process, energy consumption per operation, depending for example on the currently equipped tool head.
- **Maintenance information**. Regularly updated information about the device's runtime, tool head's attrition or other relevant influences.
- **State**. This parameter is used by the Service Orchestration to determine the device to which an object is sent.

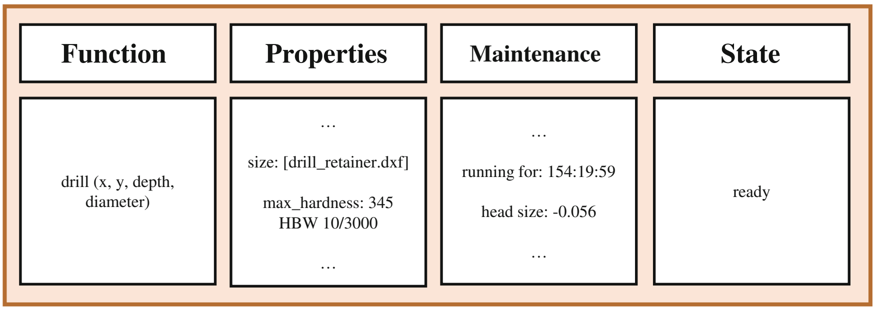


Fig. 2 Example for a possible metal drill's memory contents

Ideally, all Semantic Object Memories utilize the same ontology to model their physical object's properties. If, for example, a workpiece demands the step "drill (2311, 1963, 17, 16)" a machine has to be found which provides the function "drill (x, y, depth, diameter)", without properties prohibiting the specific parameters, such as the given size exceeding its limits.

Production floors

For the use cases, a simplified production line represented by object memories as defined above is assumed. It is divided into a floor with three drills and another floor with three milling machines. Transportation between these floors and other relevant destinations is considered to be automated.

The drills are specified as follows:

- Drill A can hold even large workpieces and drill in various depths and sizes, but only up to a certain degree of hardness.
- Drill B has the same abilities, except that it can process objects of any hardness, consuming more energy than Drill A.
- Drill C can hold and process three workpieces at once, but only if they do not exceed a certain size and with the same material restrictions as Drill A. Its energy consumption is higher than that of Drill A, but lower than using it three times.

The milling cutters are specified as follows:

- Mill A can hold even large workpieces and mill in various shapes and sizes, but only up to a certain degree of hardness.
- Mill B has the same abilities, except that it can process objects of any hardness, consuming more energy than Mill A.
- Mill C can hold and process three workpieces at once, but only if they do not exceed a certain size and with the same material restrictions as Mill A. Its energy consumption is higher than that of Mill A, but lower than using it three times.

2.1 Use Case 1: Orchestrated Production

This use case explores how an automated cyber-physical production line as described above may produce metal workpieces in a centrally orchestrated way. As defined above, arriving resources are already equipped with Semantic Object Memories containing, among other data, a set of instructions about their processing. The orchestration unit reads these instructions and devises the workpiece's further treatment.

Assume a piece of metal arrives which demands a drill operation and two different mill operations. Its shape and hardness make it suitable for Drill A which is issued to perform the operation. Meanwhile, a second and third piece have come in with drilling and milling instructions each. The second is of a harder metal and thus sent to Drill B. The third could be handled by Drill A, but the state entry in that drill's object memory marks it as busy. However, since the instructions are not explicitly ordered the workpiece can still be sent to Mill A which is idle. Once Drill A, Drill B and Mill A are finished with their operations the workpieces can be transported to their next destination. The second, harder metal piece is sent to Mill B for further processing, the first and third change places and leave the production line to their new destination after the operations are finished.

A few more resources have arrived in the meantime. One of them calls for a certain order of operations, it needs to go to the milling floor before it can be drilled. However, Mill A is busy, so the piece is sent to Mill C, together with two others which do not call for a certain order in instructions. The remaining pieces are sent to Drill C which can handle them efficiently and will swap places with the former ones later.

This way, the Service Orchestration unit checks all incoming workpieces for their specific properties and production plans and then devises their path through the workshop. The Semantic Object Memories and their contents make sure that only suitable workpieces are sent to a machine and that every workpiece is processed as soon as possible.

2.2 Use Case 2: Automated Maintenance

A cyber-physical production line is not only able to enhance the handling of resources but also of its own components. Metal processing tools are exposed to attrition and have to be maintained regularly lest they wear out or develop defects. Therefore, the Semantic Object Memory of the line's machines may hold information about its current runtime, temperature, tool attrition or other relevant aspects from which the orchestration unit can derive when it's time for examinations, tool replacements or other maintenance measures.

For example, assume that during one of its regular checks the orchestration unit finds that Drill A has been running for more than a specific amount of time since its last control, causing it to issue an examination of the machine. Its state is updated to a correspondent value, such as "inactive" or "in maintenance", for the duration of the examination and possible follow-up actions, and the timer reset afterwards. Likewise, if a temperature value exceeds a fixed threshold or a tool head becomes too short, similar steps can be taken to approach the problem at hand.

Measurements like these could also be taken in a conventional production environment, though the tracking of maintenance dates and sensor values would have to be conducted in another way, maybe even manually. However, further improvements are rendered possible only by the implementation of digital representations of all the objects in the production line, available to and readable by the orchestration unit.

For example, another possibility to discover possible production hazards is the surveillance of finished workpieces. In the described scenario the memories have a quality check entry which stores the results of an automatic quality check performed after every operation. If an object's actual condition does not match its intended one, as stated in the instructions, it fails the check. If, for example, an orchestration unit registers an unusual high amount of workpieces failing the quality check after having been processed by Mill B it might issue an additional examination of that milling cutter, as it might be damaged or tarnished before its maintenance cycle is completed.

Automated maintenance reduces the danger of material damage and inactive periods due to impaired machines. Following predetermined rules or recognizing possible dangers dynamically, a maintenance-aware Service Orchestration application can use Semantic Object Memories to keep the whole production line under surveillance and step in before any damage is caused. Its range of possibilities to detect errors is far more advanced than regular approaches, as it has a full overview of every object taking a part in the production process at any time and can monitor them individually and in relation to each other.

2.3 Use Case 3: Priority Management

Use case 1 already demonstrated how a Service Orchestration unit can make sure resources are processed with the least amount of downtime. It is entirely possible, though, to steer processing by using different criteria or even make a workpiece's treatment dependent on individual preferences of the customers ordering it.

For example, the number of units in an abandonment of production loses relevance. As the object's memories contain all necessary information no further work is needed than that which the orchestration unit performs in any case, whether the digital representation is one of hundreds or the only of its kind. A fully automated production line as presented here can thus produce large batches of one workpiece or even individual orders without the need of spending resources on devising and preparing their processing.

The priority entry in the object memories for the scenario is another option to customize production, apart from the used materials themselves and their instructions. Consider three workpieces arriving at the same time with identical memories assigned to them but for one difference: One has a "speed" priority, one an "energy" priority and one comes without a specification. The third is treated as a regular workpiece and added to the production plan as usual. The second piece has a priority to be processed efficiently in terms of energy, so the orchestration application may delay its treatment until there are enough other resources to share one of the machines that can handle three workpieces at a lower cost. The first, however, is instantly scheduled for processing, disregarding energy issues and possibly even overwriting existing schedules for pieces without the "speed" priority. For example, if Drill A is busy the piece may be sent to Drill B or Drill C instead, even though they are less energy efficient.

Priority management can thus add to a workshop's flexibility by reducing the effort needed for dynamic production to the modification of one or a few values in a Semantic Object Memory.

3 Related Work

The technologies described later in this chapter can be a valuable asset to CPS-based manufacturing and similar concepts have been introduced in the past. However, in a field where machines which were procured twenty years ago can still be considered current today and are hard to replace with newer versions innovation is difficult to achieve. Many attempts to create a common standard for similar industrial scenarios have been made but so far none was able to spread throughout the industry. Therefore, a brief overview of technologies connecting physical objects with computational elements is given here to put the presented ideas into context without a claim to completeness.

3.1 Semantic Technologies

The introduction of semantic aspects to automation is a newer development but holds a great potential for automated service orchestration. Customarily, orchestration mechanisms focus on purely syntactical approaches, such as BPEL, which without further efforts lack the possibility for applications to locate and deploy specific services based on machine-readable descriptions [8]. This ability, though, is the groundwork for the processes described in the use cases.

The "semantic gap" can be closed by different means, such as digital taxonomies or topic maps, but the use of ontologies is prevalent. These are formal collections of knowledge, describing real world phenomena, represented by concepts, instances and relations between those, in a machine-readable way. Their popularity is not only founded in their expressiveness but also the accessibility of ontology technologies. For example, the Resource Description Framework (RDF) is an XML-based formal markup language for the description of metadata. It models information in subject-predicate-object triples which can interlink and thus form a vast knowledge graph [9]. With a suitable schema file, RDF can be assigned fixed semantic rules. Even more sophisticated is the Web Ontology Language (OWL) which builds on RDF and has developed into the preferential ontology language of the Semantic Web. It can model not only taxonomies and hierarchies, but also complex logical relations and axioms [10].

Considering these are open standards, building on them as cornerstones is a widespread method to create domain-based ontologies. It is therefore no surprise that a number of such derivates also exist in the service orchestration area. For example, the OWL-based markup OWL-S for semantic web services is able to describe these services semantical, as well as modelling processes composed of them. An orchestration application can use such an ontology to discover the services it needs for a certain task and compose them according to its production goals. This has been attempted, for example, in the SOCRADES project which utilizes ontologies for the semantic discovery and orchestration of production services [11], though only in the scope of simplified setups and simulations.

AutomationML is another XML-derived open standard. It does not centre on a single device and its outputs but aims to describe all components of a complex production environment as a hierarchical structure, facilitating consistent exchange and editing of plant layout data between heterogeneous engineering tools. It thus closes existing gaps between planning, simulation, deployment, visualization and other stages, as well as the integration of different devices [12]. The plant topology is modelled in XML with pre-defined tags adding semantic information to the format. These make it possible to depict hierarchical structures of objects with attributes and relations, both physical and logical, allowing for a complete overview of all entities playing part in a production process. By utilizing existing standards such as PLCopen XML [13] or COLLADA [14] AutomationML is also able to model logic, like behaviour sequences and I/O connections, and many other information, such as kinematic or CAD data [15]. It is therefore well suited to depict most possible machines and devices employed in an automation environment and let these entities exchange data to execute their native functions based on this. However, it does not provide means adequate to model different entities, like the produced goods, for the whole of their dynamic lifecycle.

On their own, even highly developed technologies like OWL-S or AutomationML are not sufficient to provide everything necessary for the scenario depicted in the use cases in an easily accessible manner. Semantic information is important but it needs a frame to fit into an automation scenario. A solution based entirely on ontologies is not suitable for cases as mentioned above, as they require a great number of objects to have their own digital representation, filled with specific data.

3.2 Hardware Requirements

Digital representation of physical objects requires a possibility to connect both sides unequivocally. While many, though not all, current industrial machines are able to communicate digitally utilizing an increasing number of different protocols, this does not hold true for other physical objects such as resources or the products themselves. In order to link them to a digital counterpart additional hardware is necessary. Depending on purpose and frame of the object and its usage, different alternatives can be chosen.

One the one hand, microcontrollers and microcontroller-based hardware kits equip objects with their own storage and computing capacity, often even adding further features, such as peripherals, communication interfaces or installed software, up to whole operating systems, and can be added at comparatively low cost. Some of the most elaborate models include the Arduino [16], Raspberry Pi [17] and NodeMCU [18]. All of these allow for licensed or self-written software to run and make use of integrated hardware like a video controller, a real-time clock or networking adapters, thus enabling objects to become active participants in a cyber-physical system. On the other hand, this degree of versatility is often not needed and even the low prices of microcontrollers make an impact when they have to be added to thousands of objects.

Another option is object hyperlinking, analogous to digital hyperlinking the connection of physical objects with digital resources using tags and according tag reading systems. This way, a digital representation does not need to reside on the object it is mirroring but can be held, for example, on a central orchestration server, since the object in question can be identified via its link. Short range radio is one way to establish this link, requiring a transceiver to scan objects for data stored on small emitters placed inside or on them. The most common technologies in this area comprise RFID and NFC, with chips the size of a rice corn [19]. Even simpler might be a printed identifier, for example an alphanumerical ID or a graphic which can be read automatically by devices with a camera, such as barcodes, QR Codes or Data Matrices [20].

3.3 M2M Communication

Once the objects are rendered identifiable and can be connected to a digital memory some sort of interaction is required to benefit from this supplement. Communication between machines is an ongoing topic in automation and has inspired many different solutions to various problems. Despite a large number of communication standards in particular niches, on a global scale only a few technologies have grown to become widely accepted or implemented that differ in complexity, focus and supported hardware. As machine-to-machine (M2M) communication is a large area with a long history, the short view offered in the scope of this document concentrates on only a two of these, namely MTConnect and OPC UA.

MTConnect is a recent protocol standard relying mainly on Internet technology to equip manufacturing equipment with connectivity, allowing machines to communicate their current status to the outside world [21]. The data is presented in XML format and transmitted unidirectionally through information providing entities called Agents and can be retrieved via standard HTTP. This allows most applications to access the data easily, although for the machine itself oftentimes an adapter is needed which translates internal information into the data sent to the Agent. The standard is distributed under royalty-free licensing terms with the necessary schema files available online. These not only specify a common syntax for MTConnect documents but also the semantic elements, for example different categories for data items, such as “EVENT”, “SAMPLE” or “CONDITION” [22]. The fact that this data is only flowing from the device to the outside and no manipulations can be made from an external application grants security on the one hand, but also limits possible usages on the other. Unlike most production line machines which are hardly altered over many years some objects are changed regularly during their lifecycle (as demonstrated by the workpieces in the use cases) and can therefore not be represented adequately. MTConnect's strength lies in the monitoring of recurring values produced by long-running entities, not in the flexible handling of a great size of different objects.

OPC UA is the newest version of the OPC Foundation's M2M protocol specifications and adds machine-readable semantic information to the secure transport and representation of machine data, such as measured values or operation parameters [23]. After creation an OPC UA server may host any number of hierarchically sorted objects representing entities in a production line or their parts, allowing access for viewing and editing information as well as calling specific functions. A service orchestration application may connect to one or more of such servers to receive and manipulate data about the physical entities they represent. The modelling of these entities is not pre-defined and may follow any customly devised schema, as long as it is expressible in the versatile object-oriented address space provided by the OPC foundation. It is possible to view, change and delete data during runtime through various, even self-made interfaces, with the standard keeping track of changes automatically if told so. While OPC UA is one of the most flexible standards, as it can be used to represent almost any entity due to its semantic being open to custom arrangement, this flexibility also leads to increased complexity.

Many of the existing M2M standards are not interchangeable, so if aspects of two or more of them were to be combined, this leaves the options of forgoing those aspects, recreating them in another standard as good as possible or run several solutions next to each other, only using them partly. However, efforts have been made to reconcile wide-spread formats with each other, such as the OPC UA Companion Specification [24], which ensures interoperability between MTConnect and OPC UA, or even a working group between the AutomationML e.V. and the OPC Foundation aiming to connect the two [25].

3.4 Digital Object Representations

Most of the presented standards above are well suited to allow communication in an automation scenario, though they usually focus on the machines' interactions and leave out other entities. Whether it is AutomationML's fixed vocabulary which makes it hard to formulate a product's XML representation instead of a production device's, or whether it is MTConnect's unidirectionality which does not allow external changes to an object during runtime, additional representations for entities as workpieces or finished goods are often difficult to create and add. So naturally, formats have been developed to fill this gap. These start in the production environment where for example [26] or recently [27] proposed knowledge-based models for products, processes and resources alike. For our purposes, however, we aim at a model built to accompany an object's whole lifecycle from its industrial creation to its individual use by the end customers. A few approaches with similar goals in mind are Active Digital Identities, SmartProducts and the Web Thing Model.

Active Digital Identities (ADIs) are part of the EVRYTHNG Engine, an Internet of Things platform used to connect physical objects and software [28]. They are extensible data containers comprised of a number of standardized key-value pairs, such as a description, an ID or change dates, but also any conceivable custom information, like images, ingredients or production data. These containers serve to represent the concrete objects digitally in a cloud, so that other entities, especially humans, can interact with them, for example by viewing or changing their information. Access to the ADIs is simple and can be done using standard web technologies, in general HTTP requests which will yield JSON objects containing the mentioned key-value pairs, but also via a graphical user interface called "dashboard". The format's prior-ranking purpose is to connect consumer products to the web to allow, for example, object tracking or customer loyalty activities such as lotteries or other product-centred games. It is less suitable for automation, as many of its functions cannot be leveraged well in such a context or would at least require additional effort.

SmartProducts is an EU funded project which aims to connect products with a memory of proactive knowledge about the product itself, its properties and functions, but also about its surroundings, such as the location in which it is or has been, environmental data such as temperature, and personal information about its users and their preferences. The knowledge becomes proactive with its capabilities to execute certain tasks and interaction plans, for example providing media expounding its ingredients or intended usage [29]. It is enriched semantically by general and domain-specific knowledge provided through several OWL ontologies. While SmartProducts focus on the end user, providing even interfaces for multimodal access to the stored data, not only humans are possible target interactors with SmartProdcuts, but also various entities in the automation industry which might access data about the product's physical properties or processing instructions which, along with their concept of themselves as entities like the machines handling them, can play a role in goods "producing themselves" in a Service Choreography scenario.

The Web Thing Model is a result of the Web of Things Interest group, a W3C standardization activity aimed primarily at connecting physical sensors and actuators digitally by the use of web standards. Its concrete goal until August 2015 was to develop a model capable to represent a wide number of objects in the Web of Things [30]. Similar to ADIs the Web Thing Model only prescribes a few fields to provide basic information, such as change dates or ID, and beyond that allows any number of semantically not standardized data. This data can be accessed via HTTP commands and is returned as a JSON object. It is interesting to note that an actual physical object does not need to exist in order for a Web Thing to represent it, so it is also possible to create representations of places or events. While other formats technically allow this, too, the Web Thing Model is specifically designed with such instances in mind, along with more general objects. Because of this diversity in represented objects, though, only very basic definite instructions are made about its concrete structure and contents, leaving a lot to the specific implementation. No ready-to-use version exists as of yet, partly because the concrete modelling of the representations relies heavily on the scenario at hand.

3.5 Conclusion

As has been shown, a state of the art analysis yields a multitude of approaches to the problem of physical objects communicating with each other and further entities. The presented are only an excerpt of the many attempts to introduce a globally acceptable standard to the diverse world of existing protocols and data formats. Although their motivations are often similar most of these are incompatible, stemming from different origins or trying to achieve different goals.

A combination of two or more of the technologies presented could be used to implement the use cases given above, however, no single solution can fulfill all required tasks. A generally deployable standard would have to be thoroughly structured so that any entity receiving an instance of it can immediately find needed information and knows how to handle this data, yet at the same time it would need to remain open enough to express any kind of information, so it can represent all objects in a workflow in the same manner. However, flexibility and specialization generally outweigh each other. The format presented in the following section is an attempt at the reconciliation of both aspects in one model.

4 Digital Object Memories

In the following section, we present an architecture model for digital object memories intended to provide a data model to partition object memory data, a storage framework to persist and access this data (see Fig. 3), and a set of complementary tools based on [6, 31].

In the first place we have to link a physical object to its new digital representation. This is done by creating a unique identification code (ID) for each instance of an object. This ID goes along with the physical object for the entire life-cycle chain and connects the object with its digital representation. Hence, it is reasonable to attach this ID directly to the physical object to enable entities in the object's range to access the object's digital memory.

In our architectural framework a unique Uniform Resource Locator (URL) is used as memory ID. This has the advantage that such a URL not only represents the ID but also indicates the way of memory access (e.g., via a HTTP-connection through the world wide web). The memory itself can be located using two different approaches: (1) directly attached to the physical object ("on-product", e.g. with a smart label) or (2) the memory content is stored on a dedicated server ("off-product", e.g. in the web). Both ways are supported by the framework.

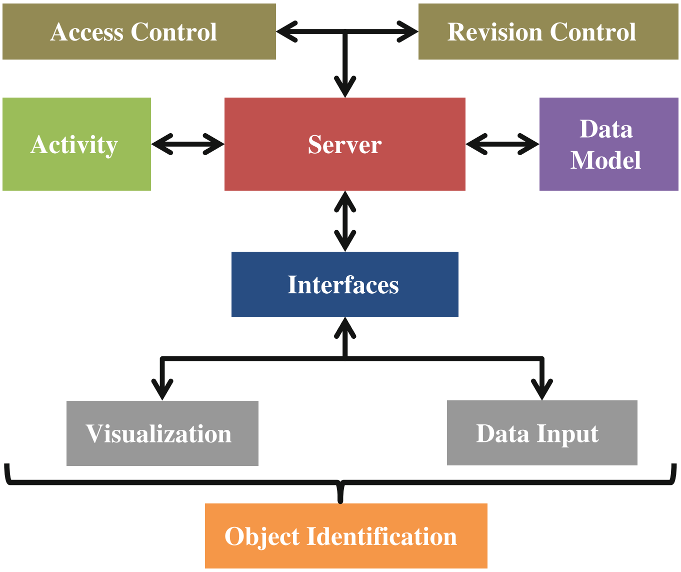


Fig. 3 Digital object model architecture

In the **off-product** case the object is a simple carrier of the read-only memory ID. This allows different solutions beginning with very cheap, printed, and optically readable 1D- or 2D-codes (e.g., barcodes, QR Codes, or Data Matrix Codes) and ranging to radio frequency identification (e.g., RFID or NFC) tags that can be accessed with dedicated readers or smart phones. However this approach requires an infrastructure to be available all the time to access the memory data.

In the **on-product** case at least a part or even the entire memory is located directly at the object, e.g., realized with small embedded systems. Hence no external infrastructure is needed to interact with the memory, with the disadvantage that such solutions are rather expensive and can be only used with upscale products. Notwithstanding the selected approach, a software component is necessary for interaction with the object's memory to parse and process the memory content.

4.1 Data Model

To fulfill the intended purpose of storing information, an object model needs a defined data model. It provides the structure which is necessary to cover the needs of very heterogeneous data created in open-loop scenarios. We propose the meta model created by the W3C Object Memory Model Incubator Group (OMM-XG) [32] and co-developed by the authors that provides a structuring element to partition the object memory information into several so-called blocks. Each block contains information of the same origin or nature (see Fig. 3) [6, 31].

Each block is a composition of two parts: the *payload* representing the stored data part and a set of *metadata* defining and describing the payload. Due to the heterogeneous data stored in object memories, access to such memories will lead to search operations (because the content is normally not known in advance) which can be performed only with the help of such metadata. The elements of the set are described in the following.

- The first attribute contains a unique identification for each block called **ID** and is represented as string.
- The **Namespace** is represented as URN and can be used to indicate that the payload has a defined content and a standardized type (e.g., "`urn:object-data:pml`" or "`urn:mycompany:millingparameters`"). This allows for direct access to the payload, if the reader supports the namespace.
- **Format** is the MIME-Type of the payload as a string (e.g., "image/png") [33].
- **Subject** contains a structure similar to a tag cloud to annotate the block payload with free text tags (e.g., "`manual`"), hierarchical text tags with a point as delimiter (e.g., "`norms.din.a4`") and ontology concepts (e.g., "`http://s.org/o.owl\#Color`").
- The **Type** attribute is a Dublin Core DCMI Type Vocabulary Type [34].
- **Creator** is a tuple of the entity that created the block and a corresponding timestamp.
- In addition, **Contributor** is a list of tuples with entities that have changed the block and the corresponding timestamps.
- **Title** contains a human-readable short title for this block and **Description** contains a longer textual description both with support for multiple languages.

The attributes ID, namespace, format, subject and type are intended for machine-to-machine (M2M) communication and are based on unique or sematic concepts. The attributes creator and contributor are used to represent a history of each block, with having a list of all contributor to this block. Finally, title and description focus on human-computer interaction (HCI) by presenting the user a human-readable text that describes the block content.

In the case that the payload has a very large size and does not fit into the memory (for example because it is located in an embedded system) or the data is redundant and used in many similar memories, the payload can be outsourced. This is done by an additional **Link** attribute that indicates the source of the payload (e.g., in the World Wide Web).

Due to the open-loop background and the focus on the shared storage space, the OMM does not provide a set of regulations for the block payload. The users are free to store the information in the way they want or in the way they have defined with other partners. However, the model includes three predefined blocks useful in several scenarios.

The **OMM-ID Block** is intended to carry all IDs of a physical object that are assigned to this object during the lifecycle. For each ID a validity timestamp or timespan can be specified.

The **OMM-Structure Block** is available to indicate relations from this object to other objects. The relations are predefined and can be combined with a validity timespan. The following relations are available: isConnectedWith, isPartOf, hasPart and isStoredIn.

Finally, **OMM-Key-Value Template** is a simple container for any data that can be reasonably represented as key-value pairs. This approach is very similar to EVRYTHNG's Active Digital Identities.

In addition, we added a proposal for another two blocks (called OMM+) extending the OMM meta model defined by the W3C XG.

The **OMM+ Semantic Block** is an extension of the OMM-Structure Block and allows the definition of arbitrary relations similar to ontology relations (e.g., provided by RDF and OWL). Each relation consists of the common triple (subject, predicate, and object) represented by uniform resource identifiers (URIs) combined with a validity statement. To indicate the physical object itself as the subject or the object the URI `urn:omm:this` is used. The predicate can be specific to a certain use case or use common RDF/OWL object relations. It is also possible to include the OMM-Structure Block relations e.g., the isConnectedWith relation by using the URI `urn:omm:structure:isConnectedWith`. This block allows the definition of semantic statements. Applications can use this information semantically (e.g., with a graph reasoner) or without a reasoner by just doing a string comparison of the triple strings.

The **OMM+-Embedded Block** is designed to integrate complete OMM-based memories into a block. This can be helpful for use cases where objects are physically integrated into larger objects. It might be the case that the memory or the ID of the integrated object cannot be reached anymore. But by copying its data to the "parent" memory, applications can still change the memory. Once the object is detached the updated memory hosted in the parent is synchronized back. To detect such blocks, a specific subject meta data attribute `primaryID.<ID of integrated object>` is used to indicate the embedded memory's ID without the need of extracting the memory itself.

4.2 Storage Infrastructure

The next tier is to apply a storage infrastructure based on the described data model [6, 31]. The foundation of all OMM-based object memory handling is the software library (libOMM) serving as reference implementation of the Object Memory Model for direct integration in Java- and .NET-based applications, to enable access to such local memories without any other functionality. On top, we extended this library to a dedicated object memory server (OMS). The server architecture is implemented on a component architecture (see Fig. 4) and can run with different

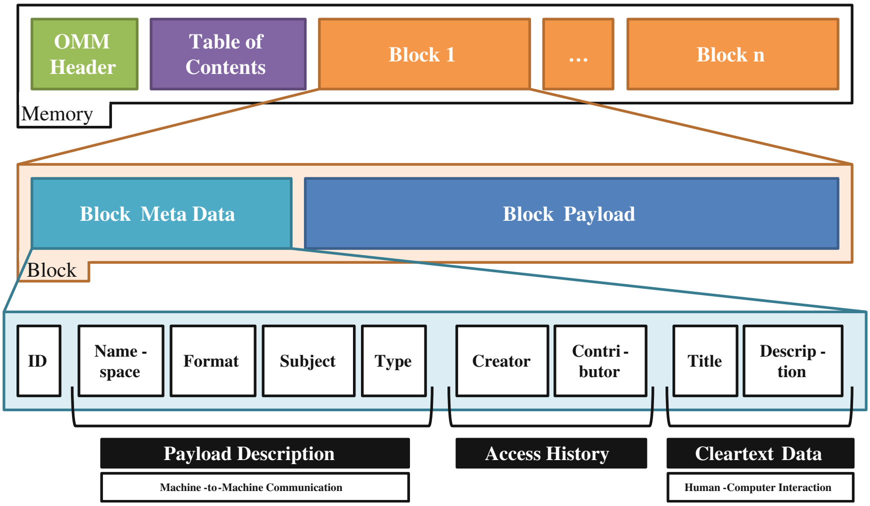


Fig. 4 Data model for object memories based on W3C object memory model incubator group (OMM-XG)

combinations of these modules depending on the intended application. Each module is implemented as servlet container. This modular approach allows the usage of the server on high-end storage systems as well as on embedded systems, bringing its extended functionality even to on-product scenarios. The communication interface and the activity module are described in the following sections. The memories are handled by the storage module that was built on top of the mentioned libOMM and can handle memories in XML and JSON representations.

An integrated revision control mechanism creates a backup of the current memory state each time the memory is altered. Users and applications can access older revisions and request a list of differences between two revisions. For debugging purposes, it is also possible to reset the memory to an earlier revision.

By using the default configuration and due to the idea of an open-loop data collector, the access to memories is not restricted generally, but the OMS is equipped with a role-based access module. The owner of a memory can restrict the two operations (read and write) for specific blocks and for entire memories. Three different approaches how to grant access to memories are available: passwords, certificates, and electronic ID cards. This process is formally defined with user U, username N_U, password P_U, certificate C_U, electronic ID E_U, memory M, block B_M in M, access right A_M and operation O in $\{O_R, O_C, O_W\}$ by the following terms:

$$F_{Name+Password}: N_U \times P_U \rightarrow U$$
$$F_{Certificate}: C_U \rightarrow U$$
$$F_{eID}: E_U \rightarrow U$$
$$F_{Access}: U \times M \times B_M \times O \times A_M \rightarrow \{True, False\}$$

The possible memory operations O are reading (O_R), creating new blocks (O_C), and modifying existing blocks (O_W). The table A_M contains all defined permissions for this memory.

A simple approach is based on username and password. This information is stored in a white-list containing all entities with access permissions. Each access is verified by the whitelist entries. The second and more secure approach utilizes digital certificates based on the ITU standard X.509 [35]. This allows for an additional way of restriction: the certificate chain mode. This mode demands that an accessing entity uses a valid certificate and this certificate must provide a valid certificate chain with respect to the root certificate of the memory. Based on this approach, we created a prototype application based on the new German ID card (nPA). This card provides an electronic identification (eID) mechanism to create a unique but anonymous and application dependent ID for each card. This can be utilized to restrict memory access to such eIDs [36, 37].

4.3 Communication Interfaces

To send or retrieve memory information from the presented storage infrastructure, three different interfaces are available (see Fig. 5) [6, 31].

For direct user interaction with raw memory data, a HTML5-based web application is available. This app presents the block-based memory structure to the user. Block payload are presented in a human readable form (only for known data types and formats) and metadata can be used for searching and filtering the memory content. It is also possible to change parts of the memory content directly in the web browser.

M2M communication can be done simultaneously in two different ways: with REST and OPC-UA. As mentioned before, each memory can be accessed with a unique URL that serves simultaneously as access point and as the object's primary ID. This URL contains the DNS name of the OMS or the corresponding IP address and the memory name. By using the RESTful interface (that focusses on web service communication) such a URL looks light this: "`http://sampleoms.org/rest/sample_memory1234`". A call to this URL retrieves a JSON-document containing information about the memory details, e.g., the storage capacity, the writing property, and the URLs of all available OMS modules. One is the storage module that handles all reading and writing operation of a memory. It can be accessed with the URL "`http://sampleoms.org/rest/sample_memory1234/st/...`". A complete description of the interface can be found in [6].

For industrial applications the OPC-UA interface is more suitable. The OMS allows for access to the memory with OPC-UA calls in the same way as it is possible with REST (see Fig. 6). This form of access, too, is made with a unique

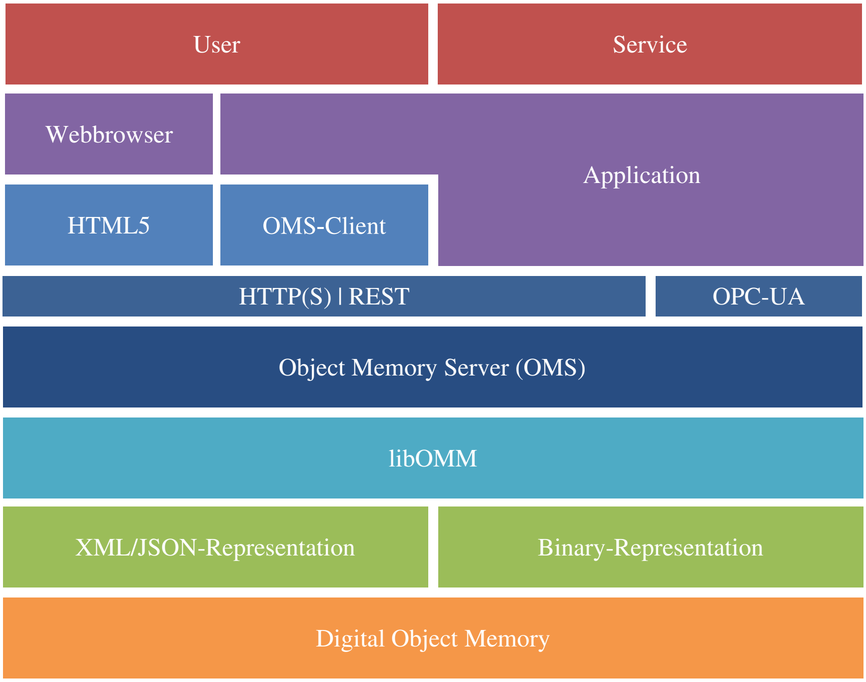


Fig. 5 Digital object model hierarchy

URL that looks like: "`opc.tcp://sampleoms.org:4711/sample_memory1234`". The block based structure is dynamically mapped to an OPC-UA data structure. All block metadata and payload can be retrieved as simple data structures. Additional calls allow for example to create new blocks or delete existing ones.

5 Semantic Service Orchestration

In this section we present a solution to equip a smart factory production line with dynamic discovery of semantic services and of orchestration of such services based on the work of Loskyll et al. [7, 8]. If a production line is producing a static set of products all the time, such a concept is not necessary. But new development demands flexible production lines that can handle different products or user specific product adaptions up to a lot size of one.

This new approach provides two new features to a production line. In the first place all devices in the line that perform a specific job are described as services. The functionality of the line is represented as a concatenation of different services forming a process (similar to a business process model). The control of such lines must handle the large variety of products. A dynamic approach can ease this job. If the devices are represented as semantic services and the objects (to be produced)

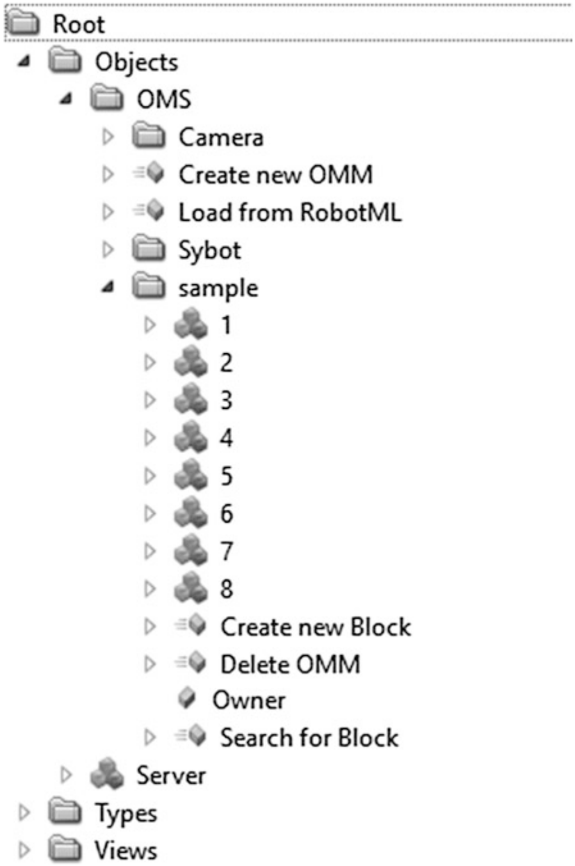


Fig. 6 UPC-UA view of digital object memory with eight blocks

provide a semantic description how to product them, a dynamic service orchestration can concatenate the best devices of the line to an optimal operation flow for all products in the second place.

Thought to the end, this approach will then able to enable scenarios as described in Sect. 2.

5.1 Service Discovery

A semantic description of services is the foundation to generate a dynamic service discovery and provides three advantages:

1. The meaning of a service definition is described in a machine-readable and machine-understandable way.

2. The relationship between components of the production and their corresponding services are defined explicitly, thus the retrieval of equal and similar services is possible.
3. Logical reasoning with the help of semantic models is possible.

This idea can be realized with commonly used semantic service description languages such as SAWSDL (WSDL-S successor) [38] and OWL-S [39]. SAWSDL adds semantics to web services by annotating certain parts of WSDL descriptions with references to semantic models. The model type is not restricted and no preconditions and effects of a service can be defined. OWL-S describes web services based on OWL ontologies by adding three types of knowledge: the ServiceProfile (what does this service do?), the ServiceModel (how does the service work?) and the ServiceGrounding (how can this service be invoked?).

To generate an efficient discovery of services and to make it reusable for different use cases in factory automation, several kinds of information are needed. Information about the field device itself covering category, type, manufacturer, and serial number of a device is the basis. In addition, the provided services of a device covering the name, capabilities and parameters of the service are also necessary. Contextual information completes the basic data set. All information sets (except the contextual data) need to be connected directly to the physical device. This can be done by attaching a digital object memory to the device. This memory contains all service related data and can be retrieved directly after the device is installed in the factory.

For the automatic registration of devices and their services, we make use of the discovery mechanism of DPWS (Device Profile for Web Services) [40], which specifies a reduced web service stack designed for the implementation of web services on resource-constrained devices such as embedded systems. These files are available by retrieving the digital object memory of the field device. As soon as a field device is available, a broadcast is sent to the production line network. The controller gets in contact with the new device, retrieves necessary information from the object memory and stores it in a device repository.

5.2 Semantic Orchestration

The presented approach of a context-based orchestration is driven by the concept of deriving the process of manufacturing based on an abstract process description. This abstract description of the product specific production process is modeled as an OWL-S *CompositeProcess*, which consists of abstract concatenated OWL-S *SimpleProcesses*. These *SimpleProcesses* could represent either basic operations (e.g., grab) or high-level functions (e.g., drilling a hole).

The concrete instance of the process is generated depending on the actual structure of the production plant and the capabilities of its field devices making use of our semantic discovery and service selection system. The orchestration of the

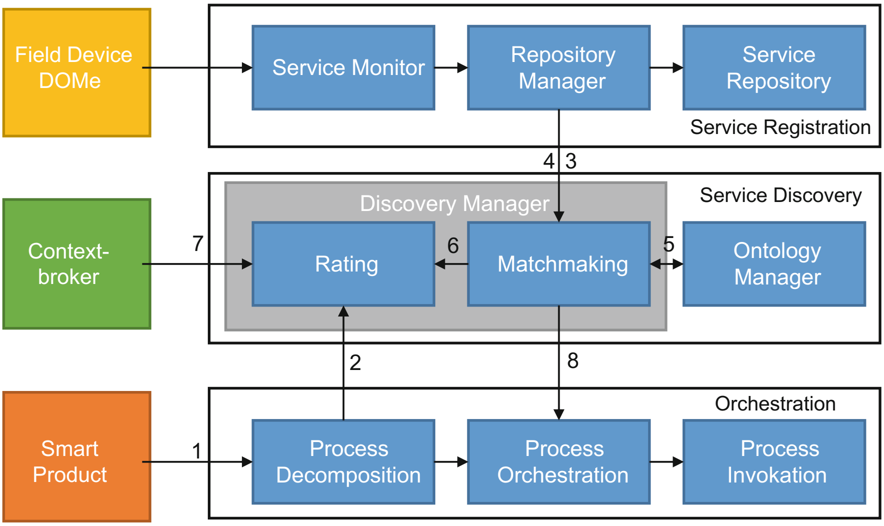


Fig. 7 Semantic orchestration workflow based on [7]

concrete process happens ad-hoc, i.e., right at the moment the product enters the production plant. Even more, the orchestration system could adapt the process at runtime (e.g., in case of a faulty component) by finding components that provide an equivalent or similar service.

Figure 7 presents a sample workflow of the given approach based on [7]. In the first step, the abstract process description of the current product (OWL-S CompositeProcess) is passed to the Process Decomposition, which decomposes the CompositeProcess into SimpleProcesses. For each SimpleProcess, the Discovery Manager is asked to find a matching AtomicProcess, i.e., a concrete web service provided by a field device or component in the plant (step 2). To this end, the Discovery Manager requests all the web services that are currently available in the plant (step 3). The Repository Manager queries the Semantic Service Repository to get back all the services including their additional information, which are sent back to the Discovery Manager (step 4). The retrieved set of OWL-S AtomicProcesses acts as input for the Matchmaker, which performs a functional matchmaking based on input/output parameters of the services (step 5). The Matchmaker generates a list of hypothetically matching services (on a functional level) and delivers it to the Context-based Rating Component (step 6). In the next step, contextual information (e.g., current state of products, machines, resource consumptions), which is queried from a Context Broker is used to influence the rating of the hypothetically matching services. In addition to contextual information, the rating process is influenced by both domain and application knowledge inferred from our different ontologies (e.g., plant ontology, equipment ontology, functional ontology). The Rating Component assigns weights to different matching criteria (e.g., provided operation, equipment category, consumed resource, Quality of Service attributes) and calculates a total score for each service. The service with the highest score is selected and sent to the Process Orchestration (step 8), which performs a re-composition of the process taking the input/output annotations of the single services but also knowledge from the function ontology and the equipment ontology into account. The procedure from step 1 to 8 is repeated until a concrete AtomicProcess is found for each SimpleProcess contained in the abstract process description. In the last step, the resulting concrete OWL-S CompositeProcess is forwarded to the OWL-S Engine, which invokes the contained services following the respective control constructs in order to control the manufacturing process for the present product in the plant.

The developed system can not only be used to realize an orchestration (ad-hoc) tailored for the manufactured product (or product variant), but also for the plug-and-play integration of new field devices. In addition, the adaption of the process in response to new and unforeseen events (e.g., device failures, changed requirements with respect to resource consumption) is possible.

6 Conclusion

In this chapter we presented work related to CPS-based manufacturing grounded on the ideas of the internet of things realized with digital semantic object memories and dynamic service orchestration.

We illustrated a comprehensive review of the current state-of-the-art work in Sect. 3, regarding data formats and data types for digital object memories and service orchestration. This review shows that there is a large variety of different formats and types available and in use. First concepts and approaches to get standardized ways of communicate product-related data between stakeholders has also been done by standardization committees, e.g. the "Referenzarchitekturmodell RAMI 4.0" [41] by the German Electrical and Electronic Manufacturers' Association.

In addition, in Sect. 4 we presented our solution for object memories that follows an open and shared common space approach. The approach has been developed prior to the RAMI model and targets the entire value chain but only subparts of the defined hierarchy levels. For direct integration purposes, the concept is realized in reference libraries and frameworks. Finally, these memories serve as data providers for intelligent objects in smart factories. We presented our solution of a dynamic service orchestration based on semantic object memory data in Sect. 5.

As a result, the envisioned ideas can be realized so that benefit for several members of the product lifecycle chain is possible. However, the large amount of different data formats and types and the challenges to create a semantic superstructure for cross-domain machine-to-machine communication thwarts the current dissemination of such technologies. In this research field more work has to be done (and is currently done) by research and industry partners to create new and to migrate existing applications to become Industrie 4.0 aware.

References

1. Heuser L, Wahlster W (eds) (2011) Internet der Dienste. Springer, Berlin
2. Tseng MM, Jiao J (2007) Mass customization. In: Salvendy G (ed) Handbook of industrial engineering: technology and operations management, 3rd edn. Wiley, New York
3. Wahlster W (2014) Semantic technologies for mass customization. In: Wahlster W, Grallert HJ, Wess S, Friedrich H, Widenka T (eds) Towards the internet of services: the THESEUS research program. Springer, Berlin
4. Müller S, Lohse W, Brecher C (2015): Integration von Produkt- und Prozessinformationen zur multidimensionalen Optimierung im Leitsystem. In: Brecher C, Kürger J, Uhlmann E, Klemm P, Lechler A, Schlechtendahl J (Hrsg.): Industrie 4.0, innovative Konzepte zur Automatisierung. (Reihe: Fortschritt-Berichte VDI: Reihe 2, Fertigungstechnik, Bd. 693). VDI-Verlag, Düsseldorf, pp 31–44
5. Kröner A, Haupert J, Hauck C, Deru M, Bergweiler S (2013) Fostering access to data collections in the internet of things. In: Proceedings of UBICOMM 2013, the seventh international conference on mobile ubiquitous computing, systems, services and technologies, Porto, IARIA, pp 65–68 304
6. Haupert J (2014) DOMeMan: a holistic framework for digital object memories. J Ambient Intell Smart Environ (JAISE) 6:507–522. IOS Press
7. Loskyll M, Dahman I, Schlick J, Schwarz M (2012) Context-based orchestration for control of resource-efficient manufacturing processes, Future Internet No. 3, MDPI, pp 737–761
8. Loskyll M, Schlick J, Hodeck S, Ollinger L, Gerber T, Pirvu B (2011) Semantic service discovery and orchestration for manufacturing processes. In: IEEE 16th conference on emerging technologies & factory automation (ETFA). IEEE, pp 1–8
9. Manola F, Miller E (eds) (2004) RDF primer. In: World wide web consortium (W3C). http://www.w3.org/TR/2004/REC-rdf-primer-20040210. Accessed 10 Nov 2015
10. Smith M, Welty C, McGuinness D (eds) (2004) OWL web ontology language guide. In: World wide web consortium (W3C). http://www.w3.org/TR/2004/REC-owl-guide-20040210. Accessed 10 Nov 2015
11. Lastra M, Delamer I (2006) Semantic web services in factory automation: fundamental insights and research roadmap. IEEE Trans Industr Inf 2:1–11. doi:10.1109/TII.2005.862144
12. Draht R, Lüder A, Peschke J, Hundt L (2008) AutomationML—the glue for seamless automation engineering. In: 13th IEEE international conference on emerging technologies and factory automation, ETFA 2008, pp 616–623
13. PLCopen (2013) TC6—XML schemes. In: PLCopen. For efficiency in automation. http://www.plcopen.org/pages/tc6_xml. Accessed 10 Nov 2015
14. Khronos Group (2015) COLLADA. 3D asset exchange schema. In: The khronos group—connecting software to silicon. https://www.khronos.org/collada. Accessed 10 Nov 2015
15. AutomationML consortium (2014) Whitepaper AutomationML Part 1—architecture and general requirements. In: AutomationML—first steps. https://www.automationml.org/o.red.c/erste-schritte.html. Accessed 10 Nov 2015
16. Arduino (2015) Arduino Genuino. https://www.arduino.cc. Accessed 10 Nov 2015
17. Raspberry Pi Foundation (2015) Raspberry Pi—teach, learn, and make with Raspberry Pi. https://www.raspberrypi.org. Accessed 10 Nov 2015
18. NodeMCU Team (2014) NodeMCU. Connect things EASY. http://www.nodemcu.com. Accessed 10 Nov 2015
19. Finkenzeller K (2015) RFID-Handbuch: Grundlagen und praktische Anwendungen von Transpondern, kontaktlosen Chipkarten und NFC. Carl Hanser Verlag GmbH & Company KG, München
20. Lenk B (2003) Handbuch der Automatischen Identifikation 1: ID-Techniken, 1D-Codes, 2D-Codes, 3D-Codes. Lenk Monika Fachbuchverlag, Kirchheim unter Teck

21. Edstrom D (2013) MTConnect: to measure is to know. Virtual Photons Electrons, LLC, Ashburn
22. Sobel W (2014) MTConnect standard. Part 1—overview and protocol. In: Standard—MTConnect. http://www.mtconnect.org/standard. Accessed 10 Nov 2015
23. Mahnke W, Leitner S, Damm M (2009) OPC unified architecture. Springer, Berlin
24. MTConnect Institute (2015) OPC UA companion specification. In: MTConnect. http://www.mtconnect.org/opc-ua-companion-specification. Accessed 10 Nov 2015
25. Henßen R, Schleipen M (2014) Interoperability between OPC UA and AutomationML. In: Disruptive innovation in manufacturing engineering towards the 4th industrial revolution. In: Proceedings of the 8th international CIRP conference on digital enterprise technology. Fraunhofer Verlag, Stuttgart
26. Raza MB, Harrison R (2011) Ontological knowledge based system for product, process and resource relationships in automotive industry. Wolfson School of Mechanical and Manufacturing Engineering; Loughborough University, Loughborough, LE11 3TU, UK. http://ceur-ws.org/Vol-748/paper3.pdf. Accessed 10 Apr 2016
27. Ferrer BR, Ahmad B, Lobov A, Vera DA, Lastra JLM, Harrison R (2015) An approach for knowledge-driven product, process and resource; mappings for assembly automation. In: IEEE international conference on automation science and engineering (CASE), Gothenburg, pp 1104–1109. doi: 10.1109/CoASE.2015.7294245
28. EVRYTHNG Ltd. (2014) Every thing connected. How web object technology is putting every physical thing on the web (white paper). In: Make products smart. https://evrythng.com/file-download/every-thing-connected-white-paper. Accessed 10 Nov 2015
29. Mühlhäuser M (2008) Smart products: an introduction. In: Mühlhäuser M, Ferscha A, Aitenbichler E (eds) Communications in computer and information science. Springer, Berlin, pp 158–164
30. Trifa V, Guinard D, Carrera D (2015) Web thing model. In: World wide web consortium (W3C). http://www.w3.org/Submission/2015/SUBM-wot-model-20150824. Accessed 10 Nov 2015
31. Kröner A, Haupert J, Deru M, Bergweiler S, Hauck C (2014) Enabling data collections for open-loop applications in the internet of things. Int J Adv Internet Technol 7: 75–85. IARIA
32. Kröner A, Haupert J, Seißler M, Kiesel B, Schennerlein B, Horn S, Schreiber D, Barthel R (2011) Object memory modeling W3C incubator group, 2011. http://www.w3.org/2005/Incubator/omm/XGR-omm-20111026/. Accessed 10 Nov 2015
33. Freed N, Borenstein N (1996) Multipurpose internet mail extensions (MIME) part two: media types. https://tools.ietf.org/html/rfc2046. Accessed 10 Nov 2015
34. Dublin Core® Metadata Initiative (2013) DCMI Metadata Terms. http://dublincore.org/documents/2012/06/14/dcmi-terms/. Accessed 10 Nov 2015
35. Cooper D, Santesson S, Farrell S, Boeyen S, Housley R, Polk W (2008) Internet X.509 public key infrastructure certificate and certificate revocation list (CRL) profile. https://tools.ietf.org/html/rfc5280. Accessed 10 Nov 2015
36. Brandherm B, Haupert J, Kröner A, Schmitz M, Lehmann F (2010) Demo: authorized access on and interaction with digital product memories. In: 8th IEEE international conference on pervasive computing and communications (PerCom). IEEE, pp 838–840
37. Brandherm B, Haupert J, Kröner A, Schmitz M, Lehmann F (2010) Roles and rights management concept with identification by electronic identity card. In: 8th IEEE international conference on pervasive computing and communications (PerCom). IEEE, pp 768–771
38. Farrell J, Lausen H (2007) Semantic annotations for WSDL and XML schema, W3C. http://www.w3.org/TR/2007/REC-sawsdl-20070828/. Accessed 10 Nov 2015
39. Martin D, Burstein M, Hobbs J, Lassila O, McDermott D, McIlraith S, Narayanan S, Paolucci M, Parsia B, Payne T, Sirin E, Srinivasan N, Sycara K (2004) OWL-S: semantic markup for web services, W3C. http://www.w3.org/Submission/OWL-S/. Accessed 10 Nov 2015

40. Nixon T, Regnier A, Driscoll D, Mensch, A (2009) Devices profile for web services (DPWS), Oasis. http://docs.oasis-open.org/ws-dd/ns/dpws/2009/01. Accessed 10 Nov 2015
41. RAMI (2015) Das Referenzarchitekturmodell RAMI 4.0 und die Industrie 4.0-Komponente, ZVEI - Zentralverband Elektrotechnik- und Elektronikindustrie e.V., Frankfurt(Main), Germany. http://www.zvei.org/Themen/Industrie40/Seiten/Das-Referenzarchitekturmodell-RAMI-40-und-die-Industrie-40-Komponente.aspx. Accessed 13 Apr 2016

Integration of a Knowledge Database and Machine Vision Within a Robot-Based CPS

Ulrich Berger, Kornelius Wächter, Alexandros Ampatzopoulos and Janny Klabuhn

1 Introduction

In today's world, goals and challenges in mechanical and plant engineering cover a wide range of research fields. The claim for global networks of plants and stations of different operators go alongside the possibility to implement extensive networks for production planning systems, energy management systems and storage systems. The pooling of resources and direct networking enable higher utilization and flexibility of plants.

The changing market conditions lead to shorter planning cycles and mass customization increases the number of variants which have to be available in decreasing lot sizes. The occurring tasks can be solved with use of modern technologies. But the conditions for integration of cyber-physical systems have not yet penetrated all production steps deep into manual processes. To implement a cyber-physical system, in particular a robot-based CPS, into manual processes the implementation of reliable knowledge databases and machine vision systems are prerequisite.

U. Berger (✉) · K. Wächter · A. Ampatzopoulos · J. Klabuhn
Brandenburg University of Technology Cottbus-Senftenberg, Chair of Automation,
Siemens-Halske-Ring 14, 03046 Cottbus, Germany
e-mail: berger.u@b-tu.de

K. Wächter
e-mail: kornelius.waechter@b-tu.de

A. Ampatzopoulos
e-mail: Alexandros.Ampatzopoulos@b-tu.de

J. Klabuhn
e-mail: janny.klabuhn@b-tu.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series
in Wireless Technology, DOI 10.1007/978-3-319-42559-7_9

2 Terms and Conditions

In order to produce the versatility of production equipment and processes in the context of the progressive dynamism of the markets, approaches like the Internet of Things (IoT) and cyber-physical systems (CPS) become necessary regarding the automation of processes. This is intended to increase productivity and efficiency of a company for a long term period [1]. CPS and IoT are integral components and enable the approaches of Industry 4.0 [2]. Therefore, various elements from the field of industry are presented in the following Sect. 4 which are used in the context of current research and implementations to make a valuable contribution to achieving efficiency gains in the industry.

Combined together, Industry 4.0, real-time networking of people, machines, objects and Information and Communication Technology (ICT) systems allow a dynamic management of complex systems [3]. The necessary CPS include software-intensive systems and devices, which represent the integration of data, services and comprehensive solutions to connect physical and digital systems to each other. CPS must be equipped with sensors that allow the adequate detection of the physical environment to analyze and interpret it in order to subsequently act purposefully with suitable actuators on the environment, so as to control the behavior of devices, objects and services. It should be ensured that the CPS can at least partially communicate via open and global information platforms. In addition, it is important to also comprise rudimentary skills such as goal-oriented adaptivity and the opportunity for self-modification based on model descriptions of their environment and their tasks [4]. Increasing communication and the control of the CPS on the Internet is called the Internet of Things (IoT) [5].

In the subsequently shown realizations of CPS technologies the focus is on the use of robots, as they are universally applicable for various manufacturing and handling tasks and possibly already with sensors on board [6]. The compound of functionalities and requirements of CPS with the ones of a robot is declared robot-based CPS in the further course. Regardless of the application, the user of the CPS has to be considered in the implementation. On the one hand, the CPS is human centered and ensures an efficient and effective human-machine interaction [7]. On the other hand, the CPS allows the interaction between robots and employees that are working simultaneously on a common working space (physical human-robot interaction [8]). Therefore, various safety aspects have to be observed in order to ensure the safety of the user and to reduce risk during usage [9]. The field of sensor technology will be dealt with separately in this chapter. Nowadays, RFID technology offers, next to the classical identification and localization, applications to carry on production-related data in a decentralized Data-on-chip principle so as to exert direct influence on production control [10]. The machine and flexible detection of objects and their position due to certain properties with the help of image processing systems is called machine vision and allows machines a detailed perception of their environment [11, 12].

3 Research Efforts

To explain the research that is done in the following use cases, it is first necessary to understand the different characteristics that belong to the main keywords.

The use cases describe the implementation of prototypes in *different surroundings*. In the companies they describe one detail of the overall process of the companies. It can either be

- an automated basic process in hazardous surroundings like the load of petrochemical liquids,
- one of several automated processes ongoing parallel in a classical production plant like the assembling process,
- an implementation to connect different automated process steps using an overall RFID system in the laundry process.

The needed *knowledge databases* differ not only in their intensity to be connected with other databases but also in their characteristics. They are always based on:

- different conditions concerning the use case,
- their update and extension possibilities,
- the number of stored parameters and attached handling routines and
- the modes to teach new items with different levels of human skills.

The state of the art is marked through a large number of progressive methods and technology which have to be chosen by the special conditions in every use case. The realization shows the current use and is open for future improvement on technical and implementation perspective.

The approach for a robot-based CPS in implementation to a research task differs depending on the requirements for the used robotic system in the use case. Handled loads, workspace, cycle times, needed for physical human-machine interaction and existing technologies within the use case scenario have influenced on the decision for either an industrial robot or a light weight robot.

The examples in the following three chapters show, how different the conditions can occur implementing a robot-based CPS within large or small and medium-sized enterprises (SME). There are fundamentally different research efforts to deal with, if the examined process does not own basic systems to rely on and is manual at the current state.

The use of modern technologies in hazardous surroundings in the frame of redesigning large scale plants includes, besides the recording of the existing technologies, also the consideration of special security requirements, being suitable for the combined deployment of technology and humans. In the chosen example of loading petrochemical liquids, long cycle times and a low variation are relevant. Due to the high forces occurring in the process an industrial robot with at least 200 kg load is scheduled. The interaction of the database is limited to the internal use of the company and might only be attached to the operational database [13].

Within the surroundings of the laundry industry the research task considers the already existing processes which are automated since decades. The approach is to handle the remaining manual processes which inhibit from increasing efficiency and process optimization. Due to the low occurring forces within the process of handling a laundry pile a light weight robot with a load of 10 kg maximum is scheduled in the second example. The cycle time in the laundry industry depends on the specific process step and is generally a very fast cycle with a very high number of parts per minute to be identified. The interaction of the database is not limited to the internal use of the laundry and should be connected to suppliers, customers and logistics [14].

In the classical surrounding of a production plant the third example shows how to develop the robot-based CPS within one of several production lines in a SME. The task is handled by hand through a human worker at the current state. The research concentrates on optimization of the assembling process with an additional quality inspection, which should take place at an early point of time in the automated assembling process. The special challenge is to implement a light weight robot in restricted space due to its working range and a high number of components to handle. The cycle time is long but the task includes a lot of components, handling routines and movements. The interaction of the database is not limited to the internal use of the SME and should be connected to the suppliers.

4 Handling Dangerous Goods Using Industrial Machine Vision

Liquid dangerous goods are transported to a large extent on railways. Increasing traffic density and volume as well as security aspects may reinforce this way of transport. While the production of petrochemical products has been largely automated, the loading of the products in tank wagons so far is mainly executed manually. The materials to be loaded, their high environmental hazard and their toxic properties create a constant threat to the health of the personnel. Therefore, in order to improve ergonomics and increase process reliability, loading automation is desired.

The materials to be loaded need special treatment when it comes to the observance of explosion protection and the environmentally sound disposal of residual quantities. In addition, the heterogeneity of the tank wagons poses the biggest obstacle to automation. The tank wagons have to comply the standards of DIN EN 12561, but come in a variety of designs. On one hand, this is due to the long lifetime of rail tank wagons of about 40 years and the associated repairs and modifications over time. On the other hand, the standards of different manufacturers can be implemented in different ways. To make matters worse, the tank wagons are provided by rental companies and therefore modifications or standardizations are difficult to perform.

Other important obstacles are the environmental conditions. The loading is carried out in the open due to the risk of explosion of the substances. Also, in addition to the explosion protection according to ATEX directive [15, 16] and meteorological conditions such as rain, snow and extreme temperatures [17] related equipment directives [18] in the development of the automation solution are observed.

4.1 Analysis of the Process Flow and the Boundary Conditions

The filling of the product in the tank wagon is completed through the round manhole cover of the tank wagon, which has a weight of about 25 kg. To open the manhole cover the personnel has to step onto the tank wagon and release the opening mechanism by hand. The opening mechanism differs depending on the design of the tank wagon and consists of a number of swivel locks and ancillary equipment. Temperature fluctuations cause pressure variations in the airtight tank wagon, which can either be low atmospheric pressure or excessive pressure. This can lead to a popping or adhesion of the dome cover.

Due to the heterogeneity of the tank wagons, an in-depth analysis was carried out in terms of the occurring types. In addition, a statistical study of process-critical characteristics was carried out. Here, especially the variants of the manhole cover of the tank wagons were considered. The result of this analysis was the differentiation of the dome cover variants in three groups shown in Fig. 1, which require each a different treatment by automating the solution.

Type A summarizes all manhole covers which conform to DIN EN 12561-6 [19]. In addition, the covers that do not comply with the standard covers, but also have four equally distributed tommy screws on the perimeter of the cover can be lifted. To open the cover type A, the knob must be loosened and swung away from the dome cover. Subsequently, the dome cover can be opened in the longitudinal direction of the tank wagon [13].

The type B tank wagon cover has a strap, which runs in parallel to the longitudinal axis of the tank wagon on the cover and is saved on the opposite side of the hinge cover by a thumb screw. After releasing the knob and folding down the tommy screw, the manhole cover can be opened [13].

In Type C, the rest of the tank wagon covers were categorized. Here you will find a variety of special versions that cannot be classified in any of the other categories due to their complicated structure. The number, type and position of the fastening covers to be solved vary greatly [13].

With 75 percent of type A it is the most common, type B accounts for 10 percent and type C fills the remaining 15 percent. Since type A and type B have the same locking mechanism and can change only the position of the thumb screw, both

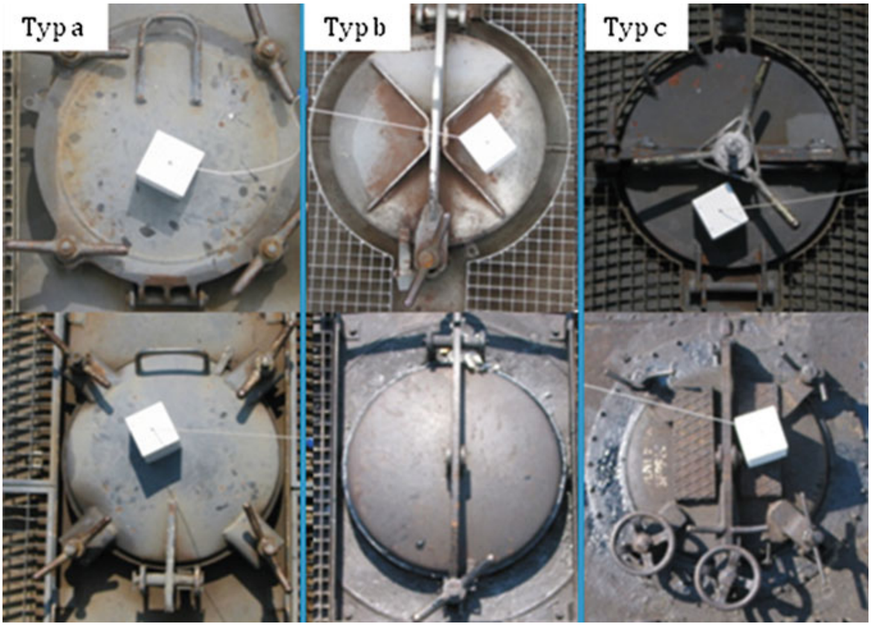


Fig. 1 Dome cover closure types

could be opened with one automated solution, which would cover at least 85 percent of all tank wagon covers [13].

4.2 Solution

The developed solution uses a depth image sensor for targeted, gradual recognition of the essential geometric parameters (e.g. tank wagon position and tommy screws alignment). Through the targeted involvement of the operator, the robustness of the system is ensured, which is influenced by the great variety of types and heterogeneity of the tank wagon and manhole cover. The integration of algorithm and parameter databases ensures that the wide spectrum of user intervention can be reduced to a minimum after a learning phase.

In the development of the gripper for the robot, mechanical solutions were preferred because of explosion protection. An air motor was used to release the tommy screws and a pneumatically driven permanent magnet to open the manhole cover. A custom adapter allows the opening of many different types of gags. The use of the latest sensor-controlled industrial robot technology increases the process safety and therefore, the labor time duration of personnel in the danger zone and the health risks are kept to a minimum. This labor and environmental protection, as well as hazardous materials requirements are now taken into account for the future.

An automated implementation is only possible with a safe and robust detection of the attack points for the robot. Thus, the positions of the tommy screws, the target of the magnetic gripper for opening the cover and the opening direction must be recognized. The investigation has shown that the tank wagon cover position does not only vary in the longitudinal axis of the tank wagon, but also in height. In order to be able to position the robot safely, all six degrees of freedom have to be determined. This can be achieved with three-dimensional or extended two-dimensional measurement methods. Pure image-processing systems cannot provide a clear picture of the six degrees of freedom due to the circumstances e.g. different paint, lighting and pollution. That the advanced two-dimensional method can cover the large measuring range in the required time is difficult to realize. To depict the entire measurement range, several tests have to be run and compositions of a three-dimensional image are required. Corresponding sensors are not available in an appropriate ATEX version [15]. Stereo camera systems can provide depth image information and can be modified in accordance with the ATEX Directive. In order to generate a depth image, a modified Microsoft Kinect sensor can be used for the development of algorithms. For the evaluation of the depth image, NI LabView Vision Development Module was used. The complex task was divided into different sections in order to achieve the required performance and accuracy. The depth image sensor has also been mounted on the gripper of the robot. So only one measuring system is required and some factors such as an elaborate calibration, controlling and positioning of the sensor are not necessary.

The Microsoft Kinect uses a patented method of the company PrimeSense, which is published under the name “Light Coding” [20]. In (computer-automation.de 2011), the process is referred to as a modification of the strip light projection. The central elements of Kinect are an RGB camera, an infrared (IR) depth sensor and a microphone array [21]. The object of the color camera is producing a two-dimensional image with a 640 × 480 pixels resolution at a frame rate of 30 Hz [22]. The Kinect depth sensor camera consists of an infrared projector and an IR camera [23]. The projector continuously transmits light with a wavelength in the infrared region (PrimeSense ™ 3D sensor). This light is invisible and harmless to humans. In front of the projector diffraction gratings are mounted, which deflect the beam and reflect it refracted into multiple beams [23]. This allows for a more defined interference pattern in which the light is emitted [23]. If the beams reach the objects and view the scene they are reflected back again (PrimeSense ™ 3D sensor). The IR camera of the depth sensor receives these reflected signals (PrimeSense ™ 3D sensor). Subsequently, the data recorded by the camera is used as a reference pattern in a pattern set relationship [23]. Based on the differences between the two patterns, different pixel depth information is determined for each pattern [23]. At the same time, in order to determine the depth image generated, the RGB camera of the Kinect is used to view a color image of the scene. This camera uses a CMOS sensor which senses differences in lightning intensity and Bayer filters that are placed in front of the diodes and the determination of the color information used [21].

In the first step the robot moves with the sensor from the tank wagon to identify the dome cover. The search area can be limited to ±2 m around the center of the tank wagon.

When the cover is found, the positions of the tommy screws are estimated starting from the center of the manhole cover. The selected areas determine individual measurements to determine the exact position and orientation. This information is required for the positioning of the end effector by the robot. The same method is also used to find a suitable point of attack for the placement of the magnet. When the positions are verified by the plant operator, a balancing of the sensor errors is carried out e.g. lens curvature, depth calibration.

4.3 Detection the Orientation of the Manhole Cover

Software was developed which enables the identification of the position of the dome cover on the tank wagon. With the aid of the depth sensor mentioned above, about 80 percent of the investigated tank wagon covers were identified unequivocally. This is achieved because of clear characteristics, such as the joint of the dome cover or the handle to open the cover. The other tank wagons do not have these characteristics. In these cases however, structures may be located above and below the dome cover to be detected. Hence, a 100 percent sensory detection of the orientation of the tank wagon is given.

These points have to be programmed separately to be distinctive for the different types of tank wagons. The algorithm cannot be identical to the other tank wagons because it is not possible due to the large number of different types. In Fig. 2, two examples are shown. In the left image, the orientation is identified by the rotary joint. Here, the depth image of a circle is formed, which is present only on one side of the cover. On the right picture a line detection algorithm was used which can detect the forms on the cover handle. The red objects show the identified locations in the image. The green area defines the search area in the searching for the distinctive property.

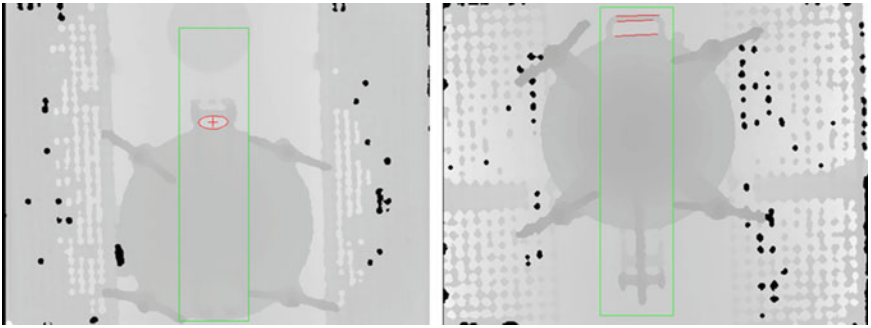


Fig. 2 Examples for identification of the orientation of the dome cover

4.4 Examination of Manhole Cover Seal

The manhole cover seal testing is an important part of the filling, since only tank wagons with a working seal may be filled. For checking purposes, environmental conditions mentioned in the previous chapter have to be followed. Tightness is achieved through various elastomer seals, which come in different versions and are located in the cover or on the tank wagon rim. The seal may be contaminated with the liquid product residues and might aggravate control. The manhole cover seal has to be free of any leaks which would cause impaired functions.

Possible error cases that need to be detected by the sensor are, for example, partially dissolved seals or missing parts. Comparable tasks in the industry are, for instance, weld control or verifying an adhesive bead. High-resolution laser line sensors are used especially to allow continuous and holistic assessment by a defined guidance on the seam. Using special software, the data is evaluated and analyzed. However, this principle cannot be applied in this case because it is difficult to adapt the sensors to the environmental conditions.

The sensor discussed in the previous section has been studied in order to be used for control purposes. The depth image may only detect objects greater than 24 mm at a measuring distance of 1 m. Flaws that are smaller cannot be reliably identified with the depth sensor. The data of the color image is very dependent on the ambient lighting conditions and therefore, it is difficult to evaluate the color and texture of the seal. A combination of both sensor color and depth information promises the best results. The position and location of the seal is detected with the depth sensor. On a later stage, the color image, which has a higher resolution, has to be searched for shade or hard transitions during the sealing process. Hard transitions and shadows are identification criteria which can provide evidence of flaws. With appropriate lighting, this effect can be enhanced. The results can then be made available to the plant personnel for evaluation. The color image of the camera with its relatively low resolution has only a limited suitability. Good results could be achieved with frames that show the close-up of the seal.

4.5 Operator Interface of the System

The operator of the plant has mainly a supervisory and controlling role. If new tank wagons are recognized by the system, the operator has to specify the constraints manually first, before the position measurements are performed. The constraints include the type of cover and the number of bolts. The measured parameters are then checked by the operator, and if necessary adjusted and saved. What already exists in the database does not need to be handled by the operator. The operator confirms the correct identification and issues the final approval.

With the two- and three-dimensional imaging of the Kinect camera, a lot of information about the tank wagons and its cover geometry can be collected.

By attaching an identification number to a unique tank wagon, the information can be stored and retrieved for every individual wagon. However, this information needs to be backed up and archived appropriately. The stored data records can be found and used when further opening steps of the wagon have to be done, e.g. to determine the number of screws and the diameter of the manhole cover and to facilitate the operator to this manual input of data.

4.6 Control and Communication Concept

Image processing technologies with computational algorithms for geometry recognition make it possible to identify the manhole cover and the swivel locks and to calculate the path of the robot to reach the desired point of application. The number of variations is handled with a database-based solution, which identifies and selects the parameterized algorithms depending on the current tank wagon. For this reason, the automation system is connected to the docking system of the overall plant system and provides the tank wagon number and information, e.g. the length over buffers.

Furthermore, the automation system is connected to the transport system of the tank wagons to reduce the search area and time for the tank wagon and its manhole cover considerably.

The sensor data collected is evaluated by means of an industrial PC which transfers the results to the user interface where the data will be verified. The cell control is affected by means of a programmable logic controller (PLC), which will

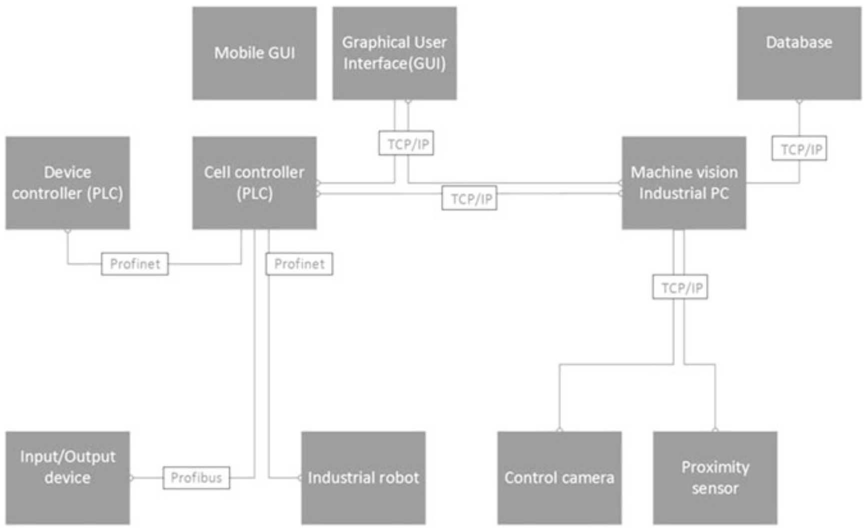


Fig. 3 Communication structure of automation

take over the management of e.g. image processing, robot and linear axis as well as air conditioning. The communication should be based on Ethernet or ProfiNet. Figure 3 shows the overview of the communication links of the system.

4.7 Conclusion

Using suitable sensors the described problem of the manhole cover opening in the hazardous zone can be automated step by step. Depth image technology combined with RGB sensor offer a great potential. It is easy to equip the system for the required safety class with this technology and in addition almost all types of closure can be detected reliably. If new variants are detected a suitable algorithm to open the cover has to be developed manually. The individual variants are then stored in a database and may in this way contribute to increasing the permanent availability of the system.

For the detection of dome sealing the depth sensor is used only for position detection due to the low reproduction accuracy. The RGB camera is used to look for close-up shadows and sharp transitions. These provide indications of voids and can be evaluated by the plant personnel on a next stage. With the help of the database and continuous development full automation can be realized. Therefore, the dangerous areas need not to be entered by personnel and as a result the risk of accidents decreases.

5 Laundry Logistics in Conjunction with RFID Systems

The influence of Industry 4.0 is analyzed in the value networks of the laundry cleaning industry. The use of available technologies in Industry 4.0 will be discussed. The dry cleaning industry serves as an example. Through the use of RFID technologies there is a significant increase in efficiency and productivity [24, 25]. RFID technology is used to identify each individual textile, which results in a piece and article recording of all laundry items. As an additional benefit for the laundry industry and the customers, a quick and easy quantification of the goods receipt is used by bulk detection [24]. In addition, the task of tracking and the upcoming amount of information lead to the research for optimization of the processes in the laundry industry.

In the laundry industry it is very common to rent textiles to hotels and other large customers. Laundries have a specific range of products e.g. linens or towels, which can be rented in individual ranges by their customers. The customer therefore purchases the service to get clean laundry delivered and returns the soiled laundry after use. Usually, hotels have an enormously high consumption of fresh laundry, therefore it is not practical to clean the laundry in house. This laundry-rental- model offers customers the advantage that they only need to operate small laundry storages and depending on the order situation they can procure different amounts of fresh laundry. In addition, there is a reduction of tied-up capital.

Figure 4 shows the supply chain of the laundry-rental-model from the perspective of a laundry facility. The goods run within the mapped supply chain from left to right. The different textiles are registered as goods and can usually be distinguished in dryer convenient laundry and flatwork. The dryer convenient laundry includes all textiles, where the residual water of the washing process is extracted in large air dryers, e.g. terry laundry towels. The flatwork can be divided into bed linen, table linen and small part laundry.

The cleaned goods are ordered in varying amounts depending on the order situation of the customer, in this case, the hotel industry. The laundries have to use sufficiently large warehouses to be able to adequately offer several customers products in good quality. Due to the high exposure during the regular cleaning process, the products need to be replaced after a very short lifespan. Moreover, it comes to a very high shrinkage of textiles within the hotel industry, which has to be compensated [24]. Therefore, new textiles are frequently ordered from wholesalers or directly from the manufacturers. These textile manufacturers are specialized in their design, confection or distribution and often work together with subcontractors [26]. The products are enhanced by other companies according to the wishes of the textile manufacturers, meaning that they are chemically or organically dyed or processed. The fabrics in turn are manufactured by weaving of different yarns. These are previously made from chemical or organic fibers in spinning mills. At the beginning of the supply chain is the fiber production, which is based on synthetic fibers, cotton, hemp or wool.

The flow of information regarding the needs of the end user is very weak due to the complex structures and therefore, decreases towards the producers. In most cases, the majority of the information reaches no further than from one segment to the next. Thus, it is very difficult for the customer to get information on where the product originates and which conditions prevailed during its production process.

Nevertheless, more information gets transferred from the manufacturer to the consumer than the other way around. The result of this low communication towards the manufacturer is called the bullwhip effect. Due to the lack of information transfer from the customer to supplier, the warehouse and production are increasing, which results in an over-reaction of the stock quantities throughout the supply chain. This process is illustrated in Fig. 4. The result of this whiplash effect is that

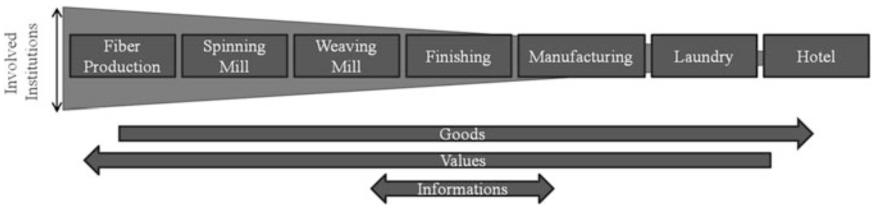


Fig. 4 Laundry supply chain

the observed demand is interpreted as the future demand. Orders are often bundled in a contract to obtain volume discounts or to gain an optimal purchase price for goods with volatile prices.

In the hotel industry, there is often a laundry storage on each floor in addition to a main storage. Since the rooms are not always fully booked, there will be fluctuations in the laundry consumption depending on the season and customer. Therefore, different amounts of products are removed from the laundry storages on each floor by the cleaning personnel. When these laundry storages are filled, there is a high demand for textiles in the main storage. Thus, the demand of several floor storages can be operated at any time, accordingly the main storage must contain enough textiles. This, in turn, is powered by the laundry with clean textiles. When there is a high demand for textiles the laundry assumes that in the future there will also be a high demand and hence, the storage size has to fit accordingly. This effect also occurs when new laundry is obtained from the manufacturers in order to compensate shrinkage and wear of the textiles or during an adaption of the storage amount. The more this process moves along the supply chain, the more extreme the demand fluctuates. The reason for this is that customers of laundries always expect an immediate satisfaction of their demand from the supplier and the supplier has to adapt its stock size accordingly. If the demand cannot be covered by the supplier, the customer can always change the supplier through the difficult competitive situation.

The only way to escape the whip effect is a continuous exchange of information along the value chain. The information about the sales volumes of the customer have to return to the supplier so the supplier is able to align his storage amounts. In order to record the storage throughput and storage volumes efficiently, each product must have its own identity and must be able to store information. A complete cross-linking of the segments of the supply chain and the information transfer by the product and the internet of things and services may lead to huge efficiency gains [27, 28].

5.1 Prospective Applications of Industry 4.0 in the Textile Industry

Industry 4.0 presents huge advantages when it comes to the laundry industry. An essential point is the improved exchange of information along the supply chain. Due to the fact that the segments of the supply chain will know at any time about the stock quantities of the cooperation partners, through the operation of a uniform cross-segment information system, they will work together more efficiently. Here, the bullwhip effect is significantly weakened and tied-up capital is reduced [27]. In addition, within the described scenarios there is a significant reduction of processing times and a clear increase in efficiency through the use of RFID technology and the automation of processes [27, 29]. The unique identification of each product allows a permanent localization and reduces shrinkage by faulty logistics processes [27, 30, 31].

Through the use of RFID transponders in the textile industry, every product can be tracked individually, which creates more transparency for the customer but even more use in the cleaning process.

5.2 Implementation of Industry 4.0 Within Industrial Laundries

In order to utilize the benefits of RFID technology within an industrial laundry, it has to be assumed first that all textiles are equipped with RFID transponders. These are used for storing information and for easy communication with the textiles. The transponders are implemented to the textiles during the production or within the industrial laundry. The special RFID tags for the usage within laundries are usually sewn to the fabric, glued or attached in the form of a label. Because of their good suitability for this application, Ultra High Frequency (UHF) transponders are very widespread. Other RFID systems are put to use, if special requirements are needed. Passive UHF transponders can store a variety of information on the product. A key point is the secure identification of the object. This includes:

- item number,
- type of the object,
- unique product number,
- manufacturer,
- owner,
- care and
- passed and maximum number of washes.

The most important characteristics of the UHF transponders spread in the laundry industry are:

- Passive transponders,
- UHF transmission frequency (865–950 MHz),
- bulk detection,
- reading distance up to six meters,
- pressure resistance in the dewatering press,
- temperature and pressure resistance at the wringing process,
- use for at least 200 washing cycles and
- purchase price for laundries and large customers under 0.50€/piece.

Currently, there are several suppliers of RFID technologies with different specification strategies. The challenge lies in identifying the examined scenario components of textile providers and connecting them with other technologies to solve the challenges in industrial laundries.

The process of picking the pieces within the laundries is visualized in Fig. 5. After the cleaning process, the clean laundry will be stored first. The textiles are placed either into the stock of the washing plant, into the storage for rental clothing or into the customer specific laundry storage. In the storages there are shelves with the stacked laundry items, where each piece is equipped with its own UHF transponder. When a customer orders a certain amount of laundry, the required stacks are automatically removed from the storages and compiled on the basis of the sales order. It is checked by means of RFID, whether all stack and related textiles are present. Based on the composition of the customer order an optimal packing scheme is determined for the laundry container via computer control. The flatwork with a high density must be placed for reasons of stability underneath the dryer convenient laundry. Before the containers leave the laundry, they are pushed through a gate with RFID readers. It will be reviewed by bulk detection, whether all components of the sales order are available.

In Fig. 6, an exemplary construction of the picking setup in the top view is presented. In this case, the laundry pile is delivered from one of the storage rooms via conveyor belt. A robot then takes the geometric properties from the RFID data

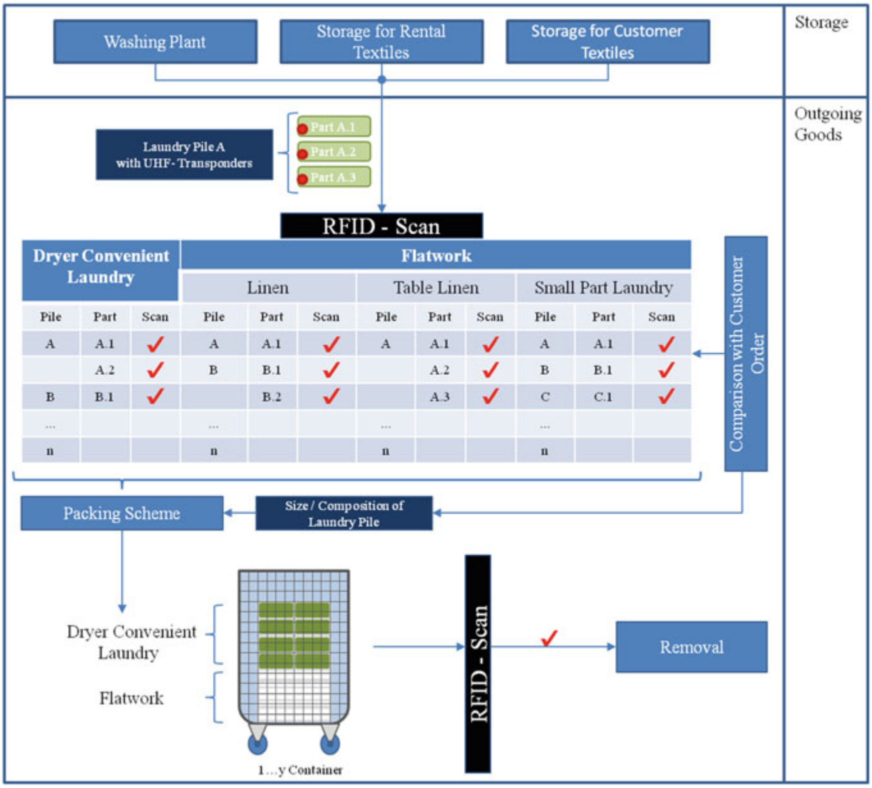


Fig. 5 Picking the textile

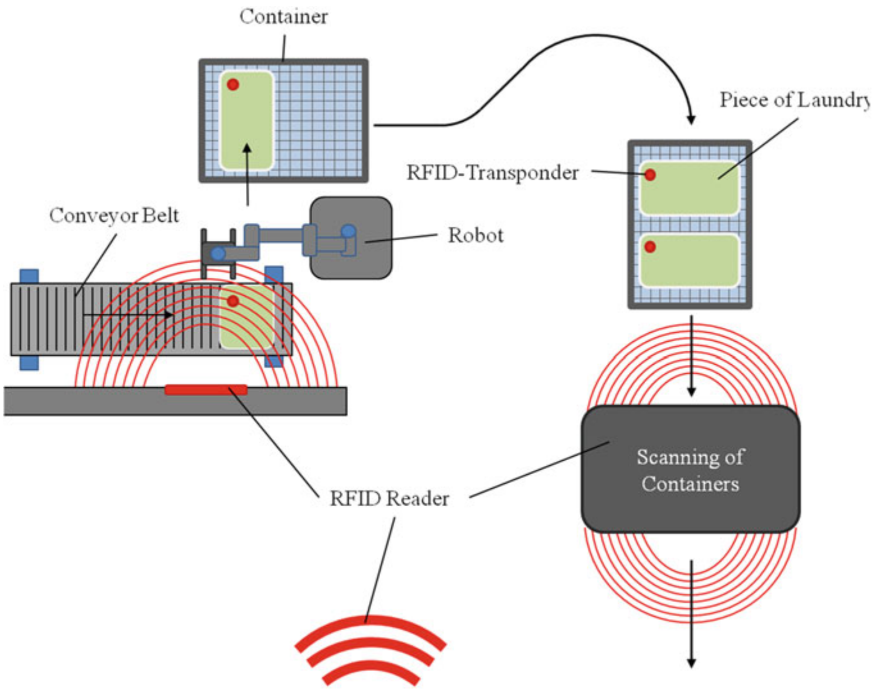


Fig. 6 Overview of the picking setup

and starts to sort laundry piles in the container. Afterwards a driverless transport system conveys this container through a RFID gate for shipping.

Once the clean laundry arrives at the customer, the delivery is checked by a mobile RFID reader for completeness. Once at the hotel, it gets stored in a sorting area on location. Once the laundry is removed or sorted from the main storage area to the hotel floors, this mobile RFID reader comes in use. This makes it possible to determine or to update the exact stock quantities at any time.

The dirty textiles are transported back to the laundry periodically and read by bulk detection. Subsequently, there is an automatic separation of the textiles depending fixed washing method, based on the information on the transponder. The machines can be individually responsive to the processing instructions and there is a direct communication between the textile tags and the machine.

5.3 Conclusion

The fourth industrial revolution offers tremendous potential to connect the segments of value chains with each other and thereby to boost the exchange of information. Accordingly, it may lead to significant increases in effectiveness along the supply chain. With the RFID-systems and their impact on the environment, there is a complete change of plans within the industry. With a more flexible production the alignment of the products can be better adjusted to the needs of customers.

For the laundry industry, the full use of RFID technology has the great advantage that the complete networking along the entire supply chain offers enormous economic potential. There are a lot of automated parts in laundry industry but also still parts of manual work, which depend on the human factor. Reaching the suitable interface between process steps effects the whole process of cleaning textiles. Higher efficiency, better working conditions and increased transparency make laundry facilities more competitive.

Action needs to be taken in conjunction with the described challenges and research in making the various technologies and information systems with suitable interfaces cooperating available. Such an orientation provides room for future innovations and potential for improved implementation of CPS in the processes of medium-sized enterprises.

6 Production, in Particular Assembly with the Help of Physical Human-Robot Interaction (PHRI)

6.1 Challenges

With the ongoing dynamical market changes, SME's are becoming increasingly dependent on using new innovations to remain competitive. This requires that companies adapt their processes to the changing market conditions such as wide variant diversity and small quantities. Larger enterprises have to change, too. Occurring problems are frequent product changes, a growing number of variations, decreasing quantities by type and smaller batch sizes. The aim is to be able to react with high flexibility on these topics through redirection of production.

6.1.1 Objectives and Approach

Based on these concrete problems with which SME's are confronted, it is the objective of this chapter to describe an example of planning and construction of a flexible robot cell. This example of a robot cell can be understood as part of an automatic production segment for the classical manufacturing environment. This concept bases on the assembly of components, which provide each a big number of variants and will be allocated into different batch sizes of the final product. The module variants of the shown assembly process differ in size, number and type of the used parts as well as they are built in a variable sequence of assembly steps.

6.1.2 Description of Process

The products in industry consist of a large variety of components. They are produced at different times and in different production processes. The task of the assembly is to convert a number of single parts in a given time to a module with higher complexity and predefined functions [32].

The described assembly process in this chapter corresponds to a typical batch production. This is a typical way of production for small and medium enterprises [33]. In the process a module is assembled, which consists of a main component and a high number of different mounting parts shown in Fig. 7.

The individual components are screws, washers, springs, rubber sealing components and a piston, which have to be inserted into the fit. Accordingly, the different shapes, materials and sizes result in different requirements for the handling of the parts in the planned automation. Currently all assembly processes are done by a worker.

After the completion of the assembly processes, a quality control is performed, which in particular checks the functionality of the fit.

In the following the planning and the implementation of a prototype for the automatically assembly cell will be explained. Therefore it is necessary to analyse the manual process first and then to identify the resulting requirements for automation subsequently.

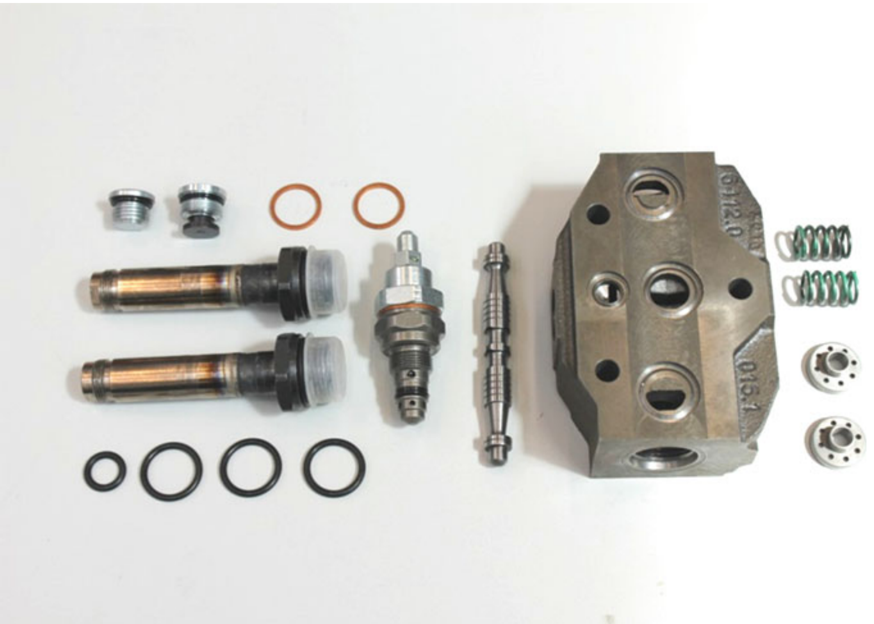


Fig. 7 Different parts of hydraulic valve system before assembly with the main component as the housing

6.2 Planning

6.2.1 Analysis

In the following the necessary assembly steps, upcoming sub problems and any desired requirements or any necessities to derive for the automation will be analyzed. The necessary sub processes for manual or automatically assembly processes are:
Isolation: forming partial quantities of a defined size from total quantity,
Positioning: moving of an object from an undefined into a defined position and
Mating: bond permanently a number of objects [34–36].

Since the different components are supplied as bulk material the location and position in the beginning is undefined. Currently, all components have to be isolated manually from the personnel. The operator takes a component out of the box whenever it is needed. Based on his competencies and cognitive flexibility a human is able to isolate and position a component in one step.

The assembling use case requires the deposition [4] and the screwing [5] of components. The deposition of the relevant screw elements is done manually. When the worker screws the component he has to place it into position by hand at first. Afterwards, he uses special tools e.g. a screwdriver for the screw elements with predefined tightening torque.

6.2.2 Sub Problems

The analysis of existing manual assembling processes show different single or combined reasons for problems. The components, delivered as bulk material, can be assembled automatically with an enormous effort because of uncertain location and position.

Tolerances within the drilled hole at the main object and the piston result in cases of lacking functionality with fits that are not adhered. That means, an increased proportion of components must be reworked. The review of is done manually by the workman but he can't check the components exactly enough. He can only use his subjective experiences and not measurable precision of perception. A repeated disassembly and assembly process accomplished from a worker can result in the same accuracy of fit with no improvement. If the quality test is situated at the end of an assembly process, it will increase the time and the costs by an inaccurate result.

6.2.3 Requirements

Analysis of former assembly processes and well known sub problems reveal several requirements for the automatic process in future. The delivered components as bulk material have to be positioned and isolated so that they can be mounted automatically. The used technology has to identify as much components as possible to isolated and mate them.

The special critical process to insert the piston has to be monitored and automated at the same time. After that mating it should be controlled that the required fit complies with essential requirements listed. The moment of the quality check for the fit should be in an earlier time in the assembly process because of the necessity to avoid elaborate remounting steps.

6.3 Applied Technologies

6.3.1 Robotics

The determined requirements call for special conditions for the selected automated solution. A compact solution should be selected which at the same time replaces many functions of a workman to reduce the number of necessary tools and facilities. An industrial robot can provide the desired flexibility and space savings. Today classical robots have a precise positioning and very high repetitive accuracy. In the given example the joining of a fit with tolerances within a few micrometres is only possible, if the conventional industrial robots would be modified.

For these sensitive and complex assembly tasks the next generation of light weight robots is suitable, in this case a KUKA LBR iiwa. Due to its structure, which is modeled based on the human arm, the KUKA LBR iiwa is similarly flexible as this, however, exceeds the precision and endurance [37]. In addition, the internal joint sensors can detect deviations between desired contact forces at the onset of the respective items and unwanted contact forces, for example in collisions of components [38].

6.3.2 Machine Vision and Data Bases

The robot is not able to grab and assemble the parts accurately, as long as it is not possible to isolate and locate the components in a supportive way for automation. Image processing systems allow the detection and localization of components within a predefined area of the cell. The recorded images can be used to generate commands for the robot to grab parts and components, which are undefined through varying position and orientation. The use of image processing systems will get more importance in production automation.

6.3.3 Pre-separation

A key point for planning the installation is the material supply. This includes the storage, deposit and providing the items at the assembly station. In conventional automation, the components are sorted and delivered in aligned form. However, the automated assembly must be carried out with the same type of material supply as in the manual assembly process. On one hand, the effort of sorting would be too large, on the other hand a change between different product variations can be carried out more quickly.

The majority of parts are provided in the form of bulk material. The as bulk material delivered items have to be isolated and positioned to be mounted automatically. This method allows processing of components with a complex geometry and forms a special case of the automated removal.

Because of the pre-separation, a 2D camera system for accurate detection of the parts orientation on the surface is sufficient. The system can reliably detect the location and rotation of the components. Herby, the components that have to be separated are taken out of the box using the robot gripper and are dropped over the separation surface. By hitting the surface each component can be separated individually. Following this, the components can be detected by using the image processing described below.

6.3.4 Robotics and Cell

When analyzing the cell and specifically the working space of the KUKA LBR iiwa it was concluded that when it is integrated into the planned cell hanging the KUKA LBR iiwa takes up space more efficient and can reach most places in the cell. The individual boxes with the separate parts can be distributed in the cell so that the required motions are minimized. The main items that are distributed are the boxes containing the various components, the separation area and the fixture to which the main body is clamped. In order for the robot to reach all necessary mounting positions of the main body, the fixture was also equipped with a swivel drive which enables the rotation of the main body together with the device by 90°.

The programming of the robot and the entire cell is to be implemented as simply and transparently as possible so that SME's who have personnel with little experience and training of robots are able to use this technology.

6.3.5 Tools Used

For grasping the various parts a 2-finger gripper was utilized which allows the robot to grab all the contours which are installed in this application example. These components are in different sizes and have different contours but most are rotationally symmetrical.

To enable the process of joining and the quality check of the fit by the actuating piston, the fingers were designed with a radius based on the curvature of the piston, visible in Fig. 8.

Subsequently, the fingers were coated with a rubber material in order to generate sufficient adhesion and some inherit compliance of the fingers at the same time. Also, the outer radius has been adjusted component-specifically, since the gripper fingers—e.g. during assembly of the coil in the main body—must be driven into the

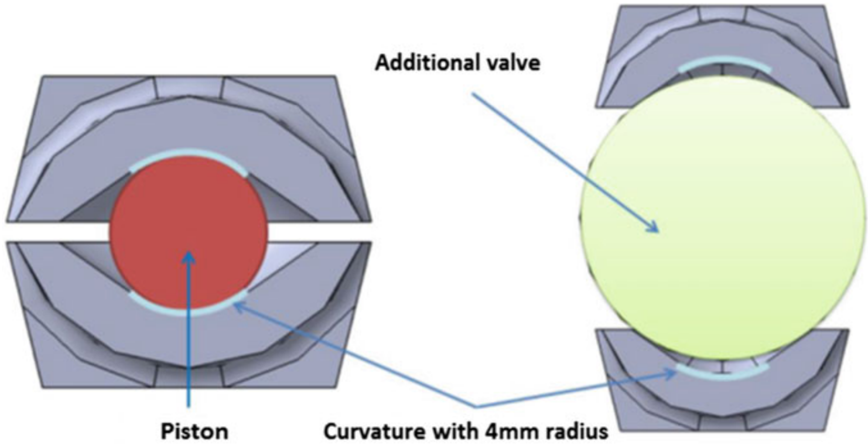


Fig. 8 Graphic gripper finger

main body in order to guarantee a secure assembly and quality testing. The gripper must not only engage different geometries but it must also realize different gripping forces, for example to grasp the sealing rings so that they do not deform and can be used by the gripper into the main body. For this reason, the electrically-driven gripper called Robotiq 2F-85 was chosen as a basis.

Furthermore, in the cell, additional tools are required to enable the overall assembly process. In this application a permanent magnet and a screwdriver are required. The permanent magnet is used for the pre-separation of the ferromagnetic components, since the 2-finger gripper can move only the central portion of the boxes due to collision in order to separate these parts. Additionally, due to the lower closing forces and indeterminate positions to the cross members it may come to the loss of these during transport to the separating surface. For this case the permanent magnet chosen was the Gaudsmit TPGC040078 because with a load capacity of 57 N it is also very well suited for the transport of the entire assembly process. In addition, it can press seal rings in the grooves provided with its flat upper surface which launched the 2-finger gripper. The permanent magnet is controlled pneumatically via a 5/2 way valve.

All threaded connections are firstly turned with the 2-finger gripper and then tightened using the wrench and the corresponding bit and torque. The change of different tools is done with the Schunk SWS-011 Tool Changing System.

6.3.6 Human Machine Interface

In order to realize modern and complex automation technologies in SME's it is important to demonstrate the simple interaction between humans and robots by means of the screwing-process. More specifically this is the case when programming new functions or new parts for robots. For the moment, important sub-functions of the robot were programmed for the testing time. It is useful to divide the program for the planned task into subprograms for differing handling routines. Depending on the skill level one will be able to get access to the programming as user, programmer, administration or robot specialist. To get the permission to enter a specific user level it is critical to check the skills of the personnel first.

Depending on the skill level of the personnel, there might be a different screen to operate the robot to choose e.g. function blocks for implementing a task solution. A worker without any special programming skills is able to select these function blocks and is led with the necessary follow-up steps in order to complete unspecified parameters. The worker has to move to the teach-in of a new position of the screwing thread, confirm with the hand-held controller and store the position. Then, an implemented subroutine automatically determines the safe insertion of the thread leads from the screw and the threaded bore. The worker can then perform a finished subroutine in which the robot automatically determines the optimum angle settings of the axes in order to maximize the screwing angle/rotation angle. Thus, a screw can be screwed into the thread by more than 360°, which protects the parts in

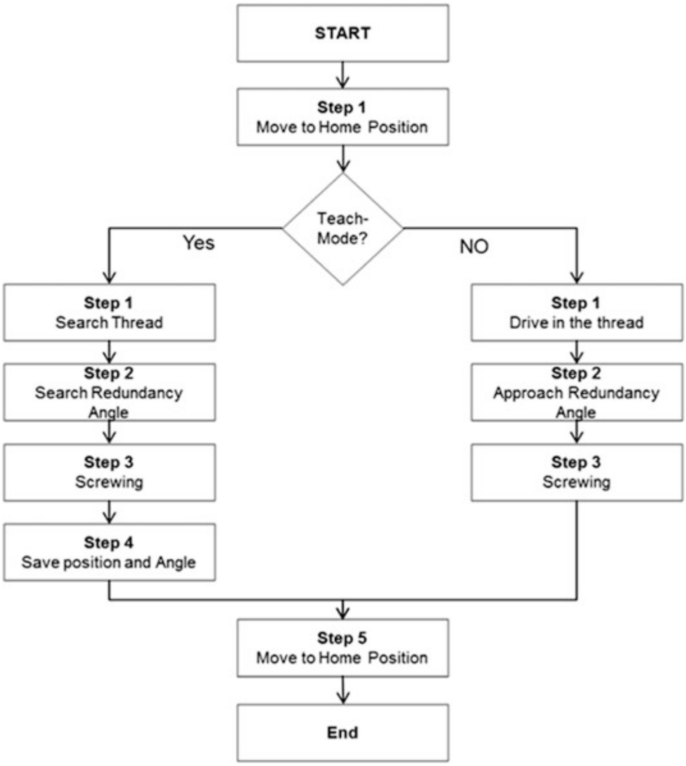


Fig. 9 Teach-mode and normal flow for implementations of a screwing process

the further handling process from falling out until a wrench is able to screw properly. The data obtained from the positions of the holes, the optimum angular positions and the data generated from the first pitch will be saved in the last step in a file for the desired process. This data then can be stored in the database in accordance with process seen in Fig. 9, *left path*.

Once the Teach-mode has been run through, the robot can determine with the final sub-programs through its sensors, the optimum angle, forces and accelerations and finally perform independently, seen in Fig. 9. In addition, the operator does not require any specific programming knowledge since these new parts and positions are already universally applicable and integrated into the finished subroutines.

6.3.7 Search Strategy

The search strategies implemented in the KUKA LBR iiwa programs, enable it to apply the tools and function blocks for a highly accurate assembly process. For example, when inserting a screw, it is first turned against the actual direction of screwing and during this process the occurring normal forces in X, Y and Z direction are monitored. In this way it can be determined when the first pitch of the screw is inserted into the first clean thread pitch of the threaded bore. Once the insertion has been detected, the screw can be cleanly screwed under the supervision of the forces. This monitoring of the forces also permits the assembly of components that have been manufactured to fit.

Here, the robot is able to approach the programmed position on the main body with the item received to then search for the hole with a loop-shaped search algorithm move with increasing amplitude and simultaneous, slight back pressure on the surface of the main body. Once the hole has been found, the respective contact force is recorded and the robot moves to the next program and performs the task. In this case the operator has the option during teaching of new components to select one of the different implemented search algorithms and does not need to customize this. Therefore, the robot can learn its positions and necessary monitored forces by itself. Nevertheless, the human worker can at any time change the parameters or teaching the parameters to change part of the steps.

6.3.8 Image Processing and Database

Among the requirements of the image processing it is important to include the localization and recognition of the individual components after the pre-separation and in the component boxes. Here, the system must be able to distinguish the different components and check if they are available in suitable quantity. When changing to other product variations, the image processing system must recognize the new components without any problems. Hence, it is critical to create an option to the effect that contains the different characteristics of component variants and image processing system provides for a change of types. This will solve a fundamentally implemented database for all components to be used in the application and handling routines.

The image processing system is adapted to the environmental conditions of the assembly process. In particular, the lighting conditions are an important factor. A reduction of stray light and changing light conditions has to be ensured. Otherwise it can be difficult to locate and identify the parts due to missing features. In addition, it must be determined at what angle the image capturing device is mounted in order to ensure an accurate possible representation of the parts.

6.4 Procedure for Developing the Solution

Firstly, image data is acquired by the components. For this reason, both the camera systems and scanning systems have to be used. This data is unprocessed and will be processed in a later stage. The first step is the pre-processing of recorded data for further processing. Through this process reliability is increased and ensures for the process to become more efficient. Possible pre-processing tasks could be the increase of brightness, the conversion to other color systems, as well as the gain or reduction of image features [39, 40]. Furthermore, morphological filters are used to alienate the components that have a complex shape [41, 42]. The characteristics of the parts become blurred and therefore convert into simple geometric shapes. The recordings of the components are high-contrast and thus the determination of characteristics by the pre-processing is simplified. It is possible for component edges to be softened or focused so that the component shape is displayed more sharply. This is particularly important in distinguishing very similar shaped components because smaller features of the parts can be highlighted. After pre-processing, segmentation of the features of the components follows. For this purpose, it is important in recordings to look for factors such as the shape, surfaces, orientations, magnitudes, colors etc. [39]. Distinctive shapes that can be searched for, are, for example, simple contours such as circles, ellipses, rectangles, and edges. The differentiation of the parts can be made by setting the radii and lengths, which ensures the distinctness of the parts due to the different geometric dimensions.

Another method that can be used is called template matching, also known as pattern recognition [39]. The basic idea of the method is the receiving and storing a reference pattern in a database. The reference sample used is then searched for within the specified image part. All parts that are consistent with the pattern are subsequently detected. After the components are detected by means of characteristics and their position on the surface or within the boxes is known, the position is transmitted to the robot. Consequently, a transformation of the coordinates of the recognized component is performed by the camera coordinate system into the robot coordinate system. After sending of the coordinates, the robot can locate the component and pick it up for the assembly process.

The entire process starting from image capturing to successfully transferring of part coordinates to the robot is made using image processing algorithms. These include pre-processing and the selected segmentation processes which are necessary for the detection and identification of the assembly parts. To determine the optimal algorithm, various trials should be carried out. For example, the changing of lighting used or the change in the height of the image capturing device and their respective impact on the image display. Here, it can be determined whether the algorithm developed in different lighting conditions and altitudes can produce the same results. In particular, reflective components can lead to problems in the use of lighting and must be considered when selecting an appropriate lighting. In addition, the evaluation of the recorded images is important for the further development of the entire system, as well as the adaptation of individual system components.

Due to the variety of options of the mounting assembly and the anticipated future need for flexibility of the cell, it is advantageous to create databases. These include a summary of the main features and characteristics of the individual components. The database in which the various data sets are stored with image examples, allows for a quick adaptation to other mounting types. For this purpose, all data must be entered at the beginning of the differentiation and identification of components and then continuously updated.

6.5 Summary

For gripping of the components in this application, the KUKA LBR iiwa will be used. To ensure the robustness and slip resistance, a 2-finger gripper is fitted with rubber linings. This ensures that parts remain at their position and do not fall out or drop. For the realization of the pre-separation a permanent magnet gripper is used on TCP of the robot. This removes the ferromagnetic components from the boxes and the robot can transport them to the separating surface and then mount them on the main component group. To assemble the parts the automatic screwdriver can be used. These can be exchanged according to the components to be mounted on the tool center point. To learn the assembly process for new component types, the robot is supported by the human personnel. Here, the teach-mode on the robot is activated, and the robot can be moved to the position necessary for assembly. For this purpose, the personnel need to activate the process of starting up the position on the human machine interface. After this mode is activated the robot is manually guided by the staff at the mounting position. This position is then stored and using the search strategies, the optimal shaft angle and angular gradients are determined for efficient installation. Upon receipt of all data, the algorithm of the mounting component is generated and stored in the database. Consequently, the stored passage can take place automatically.

In order to implement the component recognition a 2D color camera is used. This camera is mounted in a defined height above the separating surface. On this, the items are supplied in smaller quantities of surface, whereby the detection of parts is easier and the system is operating efficiently and effectively. To improve the quality of the recordings, the lighting is used. In the selection of lighting, the surface texture, reflectivity, shape and size of components must be considered. In addition, the lighting on the type of task must be selected. In this case an incident lighting is advantageous. This prevents the formation of shadow and stray light and makes it possible to produce homogeneous light conditions. Due to the better lighting, good contrast for the further processing of the images can be obtained. The subsequent localization of parts can be realized via search algorithms. The algorithms are then used to locate the rotationally symmetric components on circles, which are distinguished by a predefined radius from the system. The referenced template matching is not used here because of the increased memory requirements of the pattern, as well as the increased error rates that have arisen in the experiments. In addition, tests have shown that the use of contour detection proves to be advantageous and has a positive effect on the process cycle time. After a successful localization, the coordinates of the selected parts are sent to the robot and this can in turn initiate the subsequent assembly processes. When changing to other variants, the system is informed of the characteristics that the new components have. The geometrical dimensions of the new parts are stored in a database. When changing, the respective data set is selected from the database, so that the image processing system has access to this data. The use of the data makes it possible for the entire system to adapt to new modules series and to expand the database.

7 Research Challenges and Conclusions

The use cases describe the implementation of prototypes in *different surroundings*. They consider details of the overall process of the companies. As shown, the needed *knowledge databases* differ not only in their intensity to be connected with other databases but also in their characteristics.

The realization of the *machine vision system* in every use case, shows that future improvements on technology should be focused on improving the possibility to easily handle quality checks.

Concerning the available *robotics technologies* it can be summed up that there might be an industrial or a light weight robot for nearly every imaginable task. The current problem in this field of research is the missing of interfaces, which allow the communication to the peripheral technologies into any directions. This would support the implementation of cyber-physical systems and especially *robot-based CPS* very well.

The number of personnel being attached to robotics and somehow controlling tasks will increase within the next years. Therefore, a need for a specific level design within the robotics will be needed. The differing skill levels, complexity of tasks and specialized robotic technologies will open new dimension in the research field of *human machine interfaces*.

Moreover, the research results within automation and robotics have to be transferred into usable strategies for the small, medium-sized and large enterprises to bring economic effects to society.

References

1. Vogel-Heuser B (2013) Automation as an enabler for industry 4.0 in production on the basis of cyber physical systems. In: Vogel-Heuser B (ed): Engineering of the request until the operation
2. Kagermann H, Wahlster W, Helbig J (2013) Implementation recommendations for the future industrial project 4.0 final report of the working group industry 4.0. In: Forschungsunion im Stifterverband für die Deutsche Wissenschaft, Berlin
3. Bühler J, Bauer W (2014) Industry 4.0—Department of Economics, potential for Germany. Studie. BITKOM, Bundesverband Informationswirtschaft. Telekommunikation und Neue Medien, Berlin
4. VDI/VDE-Gesellschaft (2013) Cyber-physical systems: Chancen und Nutzen aus Sicht der Automation. Thesen und Handlungsfelder. (ed) v. VDI/VDE-Gesellschaft
5. Sendler U (2013) Industrie 4.0-Beherrschung der industriellen Komplexität mit SysLM (Systems lifecycle management). In: Baum G, Borcherding H, Broy M et al (ed) Industrie 4.0. Beherrschung der industriellen Komplexität mit SysLM, v. Sendler U. Springer, Berlin, pp 1–19
6. VDI (1990) Assembly and handling technology. Handling functions, handling equipment. Terms, definitions, symbols. Düsseldorf: VDI (VDI-Richtlinien, 2860)
7. DIN German Institute of Standardization e.V (2010a) DIN EN ISO 9241-210:2010, 01-2011: ergonomics of human-system interaction
8. Robo-Partner: Seamless human-robot cooperation for intelligent, flexible and safe operations in the assembly factories of the future (2014) 2212–8271, Elsevier B.V
9. ISO 10218-1 (2011) Robots and robotic devices—safety requirements for industrial robots—part 1: robots (ISO 10218-1:2011), German version EN ISO 10218-1:2011
10. Günthner WA, Ten Hompel M, Chisu R, Nettsträter A, Roidl M (2010) Internet of things in intralogistics. Heidelberg, Springer
11. Jähne B, Massen R, Nickolay B, Scharfenberg (1995) Technical image processing—Maschinelles Sehen. Springer, Berlin
12. Jähne B (2012) Digital image porcessing. Springer, Berlin
13. Lehmann C, Städter JP, Berger U, Keller S, Gnorski H (2013) Semi-automated handling of manhole covers for tank wagon processing using industrial robots. In: Robotic assistance technologies in industrial settings, IEEE/RSJ international conference on intelligent robots and systems (IROS), Tokyo, Japan, 3–8 November 2013
14. Wächter K, Klabuhn J, Berger U (2015) Influence of Industry 4.0 on value networks using the example of the dry cleaning industry. In: Informatik 2015, Cottbus, 28 September–02 October 2015
15. ATEX (2006) Directive 2006/42/ EC of the European parliament and of the council of 17 May 2006 on machinery and amending directive 95/16/EG
16. DIN German Institute of Standardization e.V (2010b) DIN EN ISO 12100: 2010 Safety of machinery—general principles for design—risk assessment and risk reduction
17. DIN German Institute of Standardization e.V. (1995) DIN EN 60721 Classification of environmental conditions part 3: classification of environmental parameters and their severities Sect. 4: stationary use at non-weather-protected

18. ATEX (2014) Directive 2014/34/ EU of the European parliament and of the council of 26 February 2014 harmonization of the laws of the member states concerning equipment and protective systems intended for use in potentially explosive atmospheres
19. DIN German Institute of Standardization e.V. (2011) DIN EN 12561:2011: Tank wagon part 6: manholes. Beuth Verlag GmbH; Berlin
20. PrimeSense (2013) Developers. http://www.primesense.com/developers/get-your-sensor/. Accessed 16 May 2013
21. ifixit (2010) Microsoft Kinect Teardown. http://www.ifixit.com/Teardown/Microsoft-Kinect-Teardown/4066/1. Accessed 05 Nov 2015
22. Kinectprices (2011) Kinect technical specifications. https://msdn.microsoft.com/en-us/library/jj131033.aspx. Accessed 13 Feb 2013
23. Freedman (2010) Patentnr. US 2010/0118123 A1, USA
24. Fleisch E, Mattern F (2005) The internet of things. Springer, Berlin
25. Laundry Innovation Network LIN (2015) Development of a process and the associated logistical processes for holistic RFID project in closed laundry cycle, including the processes for large customers. http://w-lin.eu/gruen/abschlussbericht_rfid_online.pdf. Accessed 13 May 2015
26. Knolle M (2006) Implementation of social standards in the supply chain of clothing companies through the establishment of networks
27. Behrendt S (2010) Development prospects of the pervasive computing. In: Siegfried Behrendt (Hg.): Integriertes Roadmapping. Springer, Berlin, pp 25–50
28. Luckett D (2006) The supply chain. In: Sammes AJ, Steventon A, Wright S (eds) Intelligent spaces. Springer (Computer Communications and Networks), London, pp 55–63
29. Anderson D, Delattre AJ (2002) Five predictions that will make you rethink your supply chain. Supply Chain Manag Rev 6(5):25–30
30. Westhuis J, Bruns R, Dunkel J, Knop W, Eichstädt T (2011) Intelligent locating of objects with complex event processing. In: Proceedings BTW—Workshops, Technische Universität Kaiserslautern, Schriftenreihe des Fachbereichs Informatik Bd. 30, pp 33–42
31. Zentrales Innovations programm Mittelstand (Central Innovation Programme SME) (ZIM) (2012) Development of a process and the associated logistical processes for holistic RFID project in a closed laundry cycle including the processes for customers. http://w-lin.eu/wp-content/uploads/2012/07/abschlussbericht_rfid_online.pdf
32. Lotter B, Wiendahl HP (2012) Mounting in the industrial production. Springer, Berlin
33. Wildemann H (2011) New installation concepts. Realization of product classification systems in small series assembly of complex products in small and medium enterprises. Technische Universität München, München
34. DIN German Institute of Standardization e.V. (2003–2009) DIN 8593-1 Manufacturing processes part 1 assembly: classification, subdivision, terms
35. DIN German Institute of Standardization e.V. (2003–2009) DIN 8593-3 Manufacturing processes part 3 pressing; classification, subdivision, terms
36. VDI-Directive 2860 (1990) Assembly and handling technology
37. Ferrara P, Kriegl V (2015) Robotics in industry and research—an overview of the robotics landscape in Austria. In: Elektrotech. Inftech, vol 132, no (4–5). Springer, Wien, pp 230–236
38. Haag M (2015) Collaborative working with robots—vision and realistic perspective. In: Alfons Botthof und Ernst Andreas Hartmann (Hg.) Zukunft der Arbeit in Industrie 4.0. Springer, Berlin, pp 59–64
39. Demant C, Streicher-Abel B, Springhoff A (2011) Industrial image processing. Springer, Berlin
40. Bässmann H (2004) Image processing ad oculos. Springer, Heidelberg
41. Beyerer J, Puente León F, Frese C (2012) Automatic visual inspection. Springer, Berlin
42. Schenk J, Rigoll G (2010) Human-machine communication. Springer, Berlin
43. Bauernhansl T, ten Hompel M, Vogel-Heuser B (2014) Industry 4.0 in production, automation and logistics. Springer Fachmedien Wiesbaden, Wiesbaden

44. Federal Ministry of Education and Research (2014) Image of the future "Industry 4.0", Bonn
45. Cyber-Physical Systems (2011) Springer, Berlin (acatech Position)
46. Herrera CD, Kannala J, Heikkilä J (2012) Joint depth and color camera calibration with distortion correction. IEEE Trans Pattern Anal Mach Intell (TPAMI) 34(10):2058–2064
47. Promotor Group (2013) Communication of research union economy—science, implementation recommendations for the future project industry 4.0
48. Embedded Systems. 1, Tagungen und Berichte, vol 3. Kassel University Press, Kassel, pp 1–4
49. Weller W (2014) On the way to the 4th industrial revolution. Technologies and their applications. Faculty I—mathematics and natural sciences

Interoperability in Smart Automation of Cyber Physical Systems

Daniel Schilberg, Max Hoffmann, Sebastian Schmitz and Tobias Meisen

1 Introduction

The digital revolution, often also referred to as Third Industrial Revolution, describes the introduction and saturation of digital technology in industrial applications. Since then the use of digital technology has drastically emerged and now touches nearly every area of life. Due to the advancement of digitalization of information and the widespread networking between arbitrarily distributed devices and systems using internet technology, the German government has originated and coined the term 'Industrie 4.0' that became known as the Fourth Industrial Revolution triggered by this advancing technologies. The outstanding success of leading corporations specialized in the use of digitalized information such as Google, Facebook, eBay and Amazon shows the potential of mass information and especially, the potential in analyzing such amounts of data. In the current days, 'Big Data' is the term loosely used to describe the exponential increase in data volumes in addition to the growth in transfer, storage and analyzing capabilities. Instead, terms like cyber-physical systems and the Internet of Things address technologies and concepts to realize the next

D. Schilberg (✉) · S. Schmitz
Institute for Robotics and Mechatronics, Bochum University of Applied Sciences, Hochschule Bochum, Lennershofstraße 140, 44801 Bochum, Germany
e-mail: daniel.schilberg@hs-bochum.de

S. Schmitz
e-mail: sebastian.schmitz@hs-bochum.de

M. Hoffmann
IMA/ZLW & IfU, RWTH Aachen University, Dennewartstraße 27, 52068 Aachen, Germany
e-mail: max.hoffmann@ima-zlw-ifu.rwth-aachen.de

T. Meisen
IMA/ZLW & IfU, RWTH Aachen University, Dennewartstraße 27, 52068 Aachen, Germany
e-mail: tobias.meisen@ima-zlw-ifu.rwth-aachen.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_10

generation of networked sensing capabilities and data propagation. These concepts form the basis to realize an 'Industrie 4.0'.

Similar terms like 'Industrial Internet' and 'Industrial IoT' (IIoT), used in other countries and industrial domains also refer to the application of these concepts and base technologies, to describe the next generation of technical systems and the market potential caused by their integration and usage in industrial application scenarios. However, providing technical systems to gather, to store and to analyze raw data is not sufficient and does not lead to the expected breakthrough in information processing. No matter how much data is available, additional context information like time, structure and location are necessary to enable the transformation of raw data into smart data and to provide the necessary information base to extract new insights by applying suitable analysis algorithms. Such transformation from raw into smart data requires context-aware analysis algorithms itself and new ways to make data accessible and interpretable for users and machines.

In a nutshell, this evolution (or revolution) leads to new sorts of human-machine inter-action and accordingly to new requirements regarding interoperability aspects. Thereby, interoperability—defined in the IEEE Standards glossary as the "ability of a system or a product to work with other systems or products without special effort on the part of the customer" [20]—is not restricted to the information exchange between technical systems. Nowadays, it has to address aspects of information exchange between human and machine as well. The human is part of the networked system and therefore the need for a reliable and robust information exchange becomes an inevitable imperative.

In this chapter we provide answers to questions raised around this emerging topic. Thereby, we rely on experience and learnings gained from many industrial and research projects we have carried out in the last decade. The non-existence of interoperability has always been one of the main obstacles to face. Instead of relying on or implementing just another silo providing solutions to current and obvious problems, infrastructures following a more holistic approach and addressing challenges as flexibility and adaptability should be decisive. In the following, we give answers to the questions:

- How can we reach interoperability on the levels machine-machine as well as human-robot?
- How is intelligence necessary to reach this kind of interoperability?
- How is the level of human-robot different and need special consideration?

The remainder of this chapter is structured as follows, after the introduction, we first focus on related work. Due to the vastness of the addressed topic, we can only give an introduction and address major trends—a detailed discussion about related work in machine-machine communication and human-robot interaction is out of the scope of this chapter. Afterwards, in Sect. 3 we apply ourselves to the machine-machine level and present applications to achieve a holistic interoperability in production networks. The human-robot interaction and especially interoperability aspects are discussed in Sect. 4. We provide two advanced use cases showing the potentials of the presented work (especially in context of machine-machine communication). Section 5 closes with a conclusion.

2 Related Work

In this section both machine-machine interaction and human-robot interaction is covered. Accordingly, the interoperability that is to be obtained needs to cover syntactical matching for machine-to-machine communication and semantical interoperability to make information understandable for both machines as part of the automation environment and human beings as deciders.

2.1 Machine-Machine Interaction

The key-feature of machine-to-machine interaction is the merging of digital and physical processes. The benefits coming from the digitalization and Big Data-related technologies can only be successfully exploited for the usage in industrial environments, if the representation of physical objects and the generation of data melt together. One approach to obtain such representation of physical objects in form of a "digital twin"—as already mentioned—are cyber-physical systems (CPS), or more precisely in the domain of production cyber-physical production systems (CPPS).

In terms of industrial production, the melting of the physical and the digital world implies that processes on the field level and information from management and planning systems from the enterprise layers make use of one coherent information model. There are several reasons, why this holistic information management needs to be established:

- Increase of flexibility in complex production systems of manufacturing companies by enabling in-process planning, reconfiguration and control.
- Decentralization of responsibilities in terms of planning and execution of production tasks to reach the aimed flexibility and customization in production processes.
- Enabling methods of learning and the usage of enhanced process knowledge directly between machines and throughout the entire information and communication supply chain of a manufacturing enterprise.

According to these goals the machine-to-machine communication on the field level has to be developed further in order to meet the requirements in terms of vertical interoperability. The focus of modern industrial protocols and data exchange interfaces lies not only on hard real-time requirements, but also on a semantic and coherent modeling of all entities that are either producing data or consuming the derived information.

There are a couple of interface standards and standardization efforts that have been carried out to meet these requirements of IT infrastructures in modern manufacturing enterprises. Starting with the development of Industrial Ethernet to more sophisticated solutions based on protocols that allow extended information modeling. The most common representatives of these semantic approaches are the Data Distribution Service for Real-Time Systems (DDS) and the OPC Unified Architecture (OPC UA) standard as a successor of the well-known and within the industrial reality well-established OLE for Process Control (OPC) standard.

2.1.1 Field Bus Systems and the Industrial Ethernet

For many decades the industrial IT and automation infrastructures have emerged from tailor-made communication solutions to generic approaches that allow reusable design patterns and infrastructure design. First attempts to generalize industrial communication was based on field bus protocols, of whom the most famous representatives are the Profibus and the Modbus initiatives mainly driven by big suppliers of industrial communication technologies. Field bus protocols work according to master-slave principles and are intended to optimize the communication in centralized automation environments. Field bus systems emphasize on the networking with programmable logic controllers (PLC) or Supervisory Control and Data Acquisition (SCADA) systems [39]. In this type of communication architecture, the production organization and execution follow a predefined, fixed logical program that has been implemented using some PLC proprietary language. The advantages of such systems include a high robustness and guaranteed deterministic real-time capabilities. The drawbacks however of these tightly-coupled systems are very limited reusability of system configuration and low scalability capabilities due to several reasons [38]:

- Strictly defined communication models—The design and implementation of field bus system communication solutions are inflexible, strongly regulated and show the same deterministic behavior as the platforms, on which they are deployed.
- Strictly defined data models—The system behavior of field bus systems is mostly to cope with rather simple data models based on I/O devices, so that the data models, which are used by field bus system, are also predefined, limited and inflexible to deal with.
- Strictly defined data types—The data types that are used for transferring information over a field bus system have to be understood by each communication participant, thus the data types employed for communicating are also rather predefined, inflexible and do not comply with any open or standard data formats.

Based on field bus protocols, advanced standards for communication in rigid automation systems have been carried out, however more focusing on an Ethernet-like communication similar to office networks and other intranets.

Networking standards like Profinet I/O, EtherNet/IP or EtherCAT were introduced to enable more distributed networking within automation environments. For example, the Profinet protocol claims to provide distributed automation solutions, managing of decentralized field devices and enable easy to manage network capabilities [21].

These Industrial Ethernet solutions extend the principle behavior of field bus protocols significantly by lifting the communication on a different layer, however the communication between the different peers of the network infrastructure is still rather rigid. The information and data models as well as data type definitions are still predefined. Industrial Ethernet still relies on tightly-coupled communication paradigms that do not comply quite well with the communication infrastructure of enterprise networks. Hence, the employment of Ethernet-enabled extended field bus protocols is still not sufficient to cope with the requirements of an 'Industrie 4.0'.

2.1.2 Data Distribution Service (DDS)

Based on rigid field bus and Industrial Ethernet solutions, more sophisticated and generic solutions for interoperability on the shop floor were carried out eventually. The DDS is such service that goes beyond standardization by providing a custom application programming interface (API) that can be used by all applications implementing the DDS protocol. More precisely, DDS is a middleware system that enables SOA-like interaction capabilities between devices on the shop floor.

The Data Distribution Service standard has been initiated by the Object Management Group (OMG) in 2004, the current version is 1.4 (released in April 2015) [33]. The motivation for DDS was to enable the passing of data between different threads of execution within distributed applications [35]. Through this type of behavior networks according to service-oriented paradigms can be established, whereby, at the same time, real-time behavior within highly decentralized applications can be guaranteed. The DDS middleware is known to be very stable and has therefore been employed especially within safety critical environments (UAVs) and by renewable and smart grid suppliers so far [40]. Other application scenarios of DDS are resource constrained environments with smaller footprints, such as resource-limited embedded devices.

The underlying technology of DDS is the Real-Time Publish-Subscribe (RTPS) wire protocol. This protocol serves as a standardized API for Ethernet based communication and follows the communication behavior of Data-Centric Publish-Subscribe (DCPS) mechanisms [35]. By providing only a standard API the DDS guarantees interoperability between DDS compliant application that are part of the same network and communicate with each other on a peer-to-peer basis. The peers of this networks (e.g. sensors and/or devices) communicate by making use of topics. The publishers provide data into the pub/sub cloud, whereas the subscribers consume data coming from all topics, they are interested in. DDS hereby adopts several mechanisms of well-established middleware and API solutions such as CORBA (Common Object Request Broker Architecture) for remote procedure calls and invocation mechanisms, JMS (Java Message Oriented Middleware API) for enabling pub-sub mechanisms and finally web services, which define service-oriented architecture approaches in distributed environments.

The unique characteristics of DDS are explicitly a low-latency data communication for highly-distributed applications, the support of "type safe" application defined data types for the creation of robust applications, support of quality of service (QoS) levels as well as dynamic discovery services. Especially the QoS level, which might be application-defined or according to a certain profile, provides ways to tailor the communication behavior of each application that is part of the communication process according to their specific needs.

Despite all of the mentioned advantages of DDS, the standard might lack the information modeling capabilities that are needed for a semantic information exchange with enterprise applications. One standard that comes along with standardized information and namespace modeling is OPC UA, which is the subject of the next section.

2.1.3 OPC Unified Architecture

In order to motivate the derivation of OPC UA, a short summary of the history in automation systems is provided. Afterwards, we present Classic OPC (OLE for Process Control), before we elaborate its successor OPC UA in more detail.

Automation systems have emerged from closed control loops on the shop floor that are merely connected through a field bus protocol. In traditional automation systems, vertical interoperability is not present—and if—then only operating from top-level systems down to the shop floor. The hierarchical organization of automation systems is depicted in Fig. 1.

As shown in the visualization, the automation pyramid reserves dedicated, separated fields for different purposes of the automation process, along with a strictly hierarchical organization of control and information flows within the production system. Information for the planning and execution of production processes are usually predefined in ERP and MES systems based on historical data or special domain knowledge in a rather manual way.

On the other side, the control flows are organized bottom-up primarily for production control and diagnosis purposes. Distributed Control Systems (DCS) and Supervisory Control and Data Acquisition (SCADA) systems use information from the field level, to guarantee the functionality of production in a determined way. However, the flow of information from the field level to the top information systems is very limited as the information that is exchanged between the low-level production systems and DCS/SCADA is restricted to information that is necessary to avoid critical system states or severe problems during operation. This data can hardly be used for a detailed analysis and optimization of the production process, because the data does not contain information about the general production state or meta-information about the sensors and actuators.

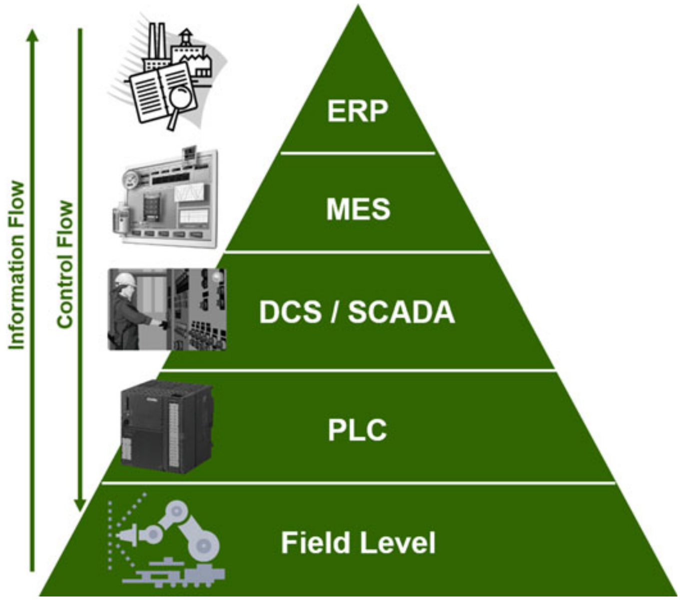


Fig. 1 Automation pyramid for the hierarchical organization of static control and information flows

In order to face the problem of separated and non-interoperable information systems along the automation pyramid, interface standards that are able to deal with both top-level and shop floor information systems need to be carried out. In a later step, the workflow for the embedding of field level devices into the ICT of the enterprise is facilitated by these standard interfaces or data protocol conventions, which enable some behavior like plug-and-play or—in terms of production processes—plug-and-produce capabilities. Ideally, these mechanisms work similar to the connection of, e.g. USB devices in computational environments.

Thus, in the beginning of the 1990s, manufacturing enterprises and major members of the automation industry attempted to carry out a standard interface that was based on the Windows NT standard as most widespread operating system throughout all companies. This OS provides the Object Linking and Embedding (OLE) technology to interconnect multiple applications on the machine. The aim of manufacturing companies was to establish similar approaches for connecting field devices with control systems of the automation layer. Thus, in 1995, a task force of big companies and automation providers like Siemens, Rockwell Automation, General Electric and ABB came up with the OLE for Process Control (OPC) standard, which uses the Distributed Component Object Model (DCOM) [4] for the linking of production facilities [22]. The central specification of this approach was manifested into the OPC Data Access (OPC DA) specification and has been published in 1996 [24]. By standardizing the access to information in automation systems, it is possible to embed driver or interface specific information directly into the field devices, similar to a USB stick for computers. This approach enables plug-and-produce capabilities without a manual configuration/integration of each device.

The basic communication functionalities of OPC DA are based on the traditional server-client principles. The OPC Server represents data from real-world objects of the field level or from other data generating devices. OPC Clients are able to connect to these servers, to subscribe to certain information and to trigger specific services. In terms of building an OPC network, the OPC Client creates an instance of the server. This server represents a single device or a group of data sources that represent concrete objects, e.g. sensors or actuators in an automation system. These items can be grouped to OPC Groups for similar items. The OPC Server represents these items and provides access to single or aggregated data to the OPC Client that is connected. OPC Servers are able to store information and to provide services to the data, whereas OPC Clients read and redistribute this information. To enable the composition of complex networks with many servers and clients, every OPC Client can also function as an OPC Server, i.e. has reading functionalities and provides access to the data for other clients.

Due to the success of this approach, industrial users demanded for additional functionalities for the OPC standard which are necessary for a full migration of the automation system in a widespread production network. These demands resulted in two additional specifications, the OPC Alarms & Events (OPC A&E) and the OPC Historical Data Access (OPC HDA) specification released in 1999 and 2001. The OPC A&E specification provides services to trigger real-time actions based on events or critical system states that could be harmful to the process or to other systems, whereas the OPC HDA specification provides functionalities to access data from previous processes for data acquisition and diagnostic purposes.

The success of OPC in the 1990s relied on the high performance of the DCOM technology and the robustness of the resulting automation system [24]. However, the composition of OPC-based automation systems is still rather rigid and hierarchical. Another drawback of OPC Classic is the lack of communication capabilities in networks as the DCOM standard cannot be properly configured to work with firewalls, which can be a problem especially in huge, distributed network environments. As a result, the OPC consortium figured out ways to communicate by making use of internet based approaches, eventually publishing the OPC XML-DA standard that enables propagation and integration of OPC-based information via web services [22]. However, OPC XML-DA turned out to be comparatively slow, mainly due to the high overhead of information integrated into each XML message for each piece of information that is sent between OPC servers and clients.

Due to the resulting interoperability problems of classical OPC solutions as well as the dependency on Windows-based systems and bad configurability characteristics in terms of industrial networks, the demand for a modernized version of OPC rose up from the industry. Consequently, the OPC Unified Architecture (OPC UA) standard had been carried out, which attempts to combine all specifications of the OPC Classic standard. Furthermore, it extends the functionalities while enabling platform independence and interoperability through web services [12].

One major functionality of OPC UA is the possibility to configure lightweight and quick connections over the internet. As OPC UA is not solely based on DCOM, the communication can be performed over network borders and firewalls. This approach solves major security problems and allows data exchange between distributed systems, as an encapsulation of system devices into a dedicated network—unlike to OPC Classic—is not necessary when using OPC UA [23].

Following as successor of the widely accepted OPC (Classic) interface standard, OPC UA is present in the industrial reality of many enterprises. Towards an information technological enhancement in terms of 'Industrie 4.0', OPC UA is likely to become the new de facto standard for the implementation and realization of scenarios in the scope of the Fourth Industrial Revolution. The OPC Foundation has already performed the first step towards 'Industrie 4.0' by integrating an extensible, scalable meta model into the standardization of OPC Unified Architecture. According to Leitner and Mahnke, the OPC UA already provides a service-oriented concept: "By defining abstract services, OPC UA provides a service-oriented architecture (SOA) for industrial applications—from factory floor devices to enterprise applications. OPC UA integrates the different flavors of the former OPC specifications into a unified address space accessible with a single set of services" [23]. The core of the service-oriented character of OPC UA lies in the definition of an abstract set of services (Part 4) [19], which can be mapped using different technologies for the communication and exchange of information (Part 6). The key feature of OPC UA is the separation of its communication stack and APIs from the information modeling. The communication stack in OPC UA is used on both server and client side to encode and decode message requests and responses [23]. The mapping ensures interoperability between the different communication stacks of servers and clients. The API on both sides can differ from each other as long as the technology mapping of request and response messages match. This introduces fundamental changes in comparison with the OPC (Classic) programming concepts. Through the new full SOA-based distributed networking paradigm, OPC UA delivers a genuine platform independence. This allows a new approach to information and service integration [17].

The presented concepts and communication paradigms that are realized by OPC UA make OPC UA to the next-level interface standard for industrial automation and beyond. The separation of information modeling and transport functionality is exactly in the scope of data exchange mechanisms that rely on information modeling on the other hand and free choice of the transport mechanisms for performance purposes on the other hand.

2.2 Human-Robot-Interaction

Over the last decade the research area of robotics, and especially the part of human-robot-interaction, has made a great leap forward. It is the aim of many projects to increase the capabilities of robotic-systems. There is a huge range of varying approaches to achieve the goal to make robots smart—according to their intentions and behavior. Approaches can be assigned to different main objectives. These objectives are related to the robots' behavior and to the information models that are used to match the behavior for a certain situation. To build up the above-mentioned robotic information model we still have to combine some existing technologies and develop them. First we take a closer look on the most important technologies for the creation and usage of the robotic information model. Second we focus on related work regarding the behavior and intention of robotic interaction more closely.

The core question regarding robotic interactions is how to connect the movements of joints to complex operations and how to explicate intentions via movements to the human. A robotic information model should fulfill these operations. To build this kind of information model we have to use methods from information retrieval and the semantic web. Furthermore, we have to look at existing information models in other application domains.

Information retrieval (IR) is marked by vague requests and insecure knowledge. Belkin et al. [3] describe the retrieval strategy in the following way: "The goal of information retrieval is to resolve those anomalies in a person's state of knowledge, which induced him or her to seek information from literature. Our approach is to select search strategies with explicit reference to characteristics of the enquirer's ASK (anomalous states of knowledge) structure." The challenge of intentions is that the state of knowledge concerning the intention of a person is uncertain and that makes it complicated to derivate a decision. Manning et al. [25] gives a comprehensive introduction to IR.

However, IR is the way to process the information, but to do so there is a need for representation of the robot's capabilities, the robot's tasks, the intentions and their relationships. Therefore, the method must enable the possibilities to discover and exploit domain specific knowledge. The concept is based upon ontologies and planning algorithms from artificial intelligence. According to Gruber [16] an ontology is an explicit specification of a conceptualization [16]. The used ontologies are expressed in OWL, which is the ontology language recommended by the W3C [42]. The matching between capabilities and intuition is related to the semantics of concepts. These concepts are implemented in diverse projects, but the most important project for our implementation is the Cluster of Excellence "Integrative production technologies for high wage countries". During the project a framework was developed that facilitates the semantic integration and analysis of measurement and enterprise data according to real-time requirements. Semantic technologies are used to encode the meaning of the data from the application code. Herewith the data is automatically annotated using terms and concepts taken from the application domain. Furthermore, a semantic integration and transformation process is facilitated. Thus subsequent integration and, most importantly, analysis processes can take advantage of these terms and concepts using specialized analysis algorithms [29, 31, 32].

Besides the core technologies used to create a robot information model, behavior and intention as psychological topics have to be taken into account. This should enable a robotic system to explicit its intention and to be able to derivate the intention of a human being out of its behavior. There are a lot of empiric studies concerning behavior and intention. These studies will be the base for the connection between human behavior and human intentions that are represented in the ontology.

Our assumption is that the human-robot interaction will be purposeful, so that unplanned behavior is ignored while the following theories are focused. The theory of reasoned action (TRA) [15] and the theory of planned behavior (TPB) [1] are the psychological background for the derivation of the intention to reach a certain goal through the behavior. The psychological studies will also help to predict the behavior. The aim of this study is to show “that discovery of the role of intentions depends on the statistical power of test procedures, the reliability of measures of intentions, and the nature of the processes intervening between intentions and behavior” [2]. According to Bagozzi et al. [2] and the study by Budden and Sagarin [5]:

> Many individuals intend to exercise, but fail to link this intention to behavior. The present study examined the impact of an implementation intention intervention (i.e., instructions to form specific if–then plans) on an exercise intention–behavior relationship among working adults who varied in reported occupational stress levels. Results indicated that implementation intentions backfired, such that participants who did not form an implementation intention exercised significantly more than participants who formed an implementation intention.

That focus on the implementation of intentions and the influence of boundary conditions e.g. like stress. We want to sort out side effects of human behavior that are related to “bad habits” of the interacting humans. We presuppose a mindfulness when it comes to the relation between behavior and intention like it is examined by Chatzisarantis and Hagger [7]: “These findings suggest that mindfulness is a useful construct that helps understand the intention-behavior relationship within the theory of planned behavior.”

The robotic information model is the “brain” of the interaction system that is able to derivate intentions out of the behavior based on the before mentioned psychological knowledge. The next step is to take a look at the behavior of the robot. A human coworker should be able to derive the objective of each operation from the robot’s actions. Therefore, the robot must act as the human expected it. The movements of an industrial robot are very efficient but not always understandable for the human user. To achieve that the robot moves exactly as the user expects, we have to add kinematic constraints to the information model. The research of Dragan et al. takes a look at misleading robot motion and lead to the result that the robot should hide its intention [10]. The conclusions of this study are very interesting for the definition of the bandwidth of acceptable movements out of the human perspective that are implemented in the information model. In another work of the authors one predicts an initial state due to previous experiences and use them for trajectory optimization [9]. The mapping from machine movements to human movements and vice versa is a basal function of the information model with the goal to enable the human to predict what the robot will do and in reverse [11]. The publications [13, 14, 27] refer to a planning algorithm that is used to self-optimize the problem-solving strategy of a robotic system. This algorithm will be used to adopt the robot system to a known situation that is directly derived from an unknown situation. Vieritz et al. [41] show a human centered design approach for the development of automation systems that is used in this work.

3 Holistic Interoperability in Production Networks

3.1 Interoperability on Machine-Machine-Level Using OPC UA and Semantic Technologies

The goal of the developments regarding the design of automation infrastructures in terms of SOA approaches, is to interconnect yet hierarchically organized systems of the still existing automation pyramid through vertical, interoperable interface standards (see Fig. 2).

In the figure, the generation and usage of information is still separated from each other through the different vertical layers of the pyramid. However, the presented approach goes beyond, a unique information flow from top to bottom by enabling a semantic description and annotation of data that is generated on the field level, respectively the shop floor. To reach the goal as shown in the picture, the information from all levels of the production is enriched with meta-data as well as production information from a broader perspective. If this information can be aggregated according to a uniform semantic description,—in terms of an integrated information model—the aggregated information can be directly used at the

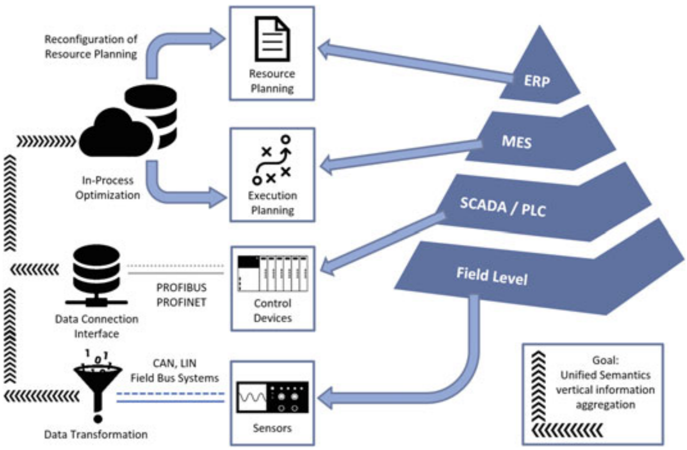


Fig. 2 Interoperability between the layers of the automation pyramid, interconnected through semantic interface standards

top-level of the production system for reconfiguring and optimizing production in-process according to the current and specific needs.

This interoperability approach based on OPC UA is still able to assure horizontal integration of field devices and other production near components by device specific drivers, but also enables vertical integration between different layers of the automation hierarchy. This includes not only connectivity between the shop floor level and production control layers, but also an integration into MES and ERP systems (as part of the own company or other related enterprises). The communication within OPC UA infrastructures can be realized using XML based web services or a UA Native approach that is based on a binary protocol [17]. The intention of these two information exchange methods is to cover the wide area of applications, which are addressed by OPC UA: data-intensive and extensible machine diagnostics applications (complex and detailed web service definitions) or real-time data of sensitive application for critical tasks close to the field level (UA binary approach). The XML based web services approach is a SOA compatible protocol confining to last WS Security specifications [19] (WS-Security, WS-Trust, WS-Secure Conversation, Public Key Cryptography Standards, Digital Signature Standard, Advanced Encryption Standard, Elliptic Curves DSA). The mapping using web services intends to enable an easy vertical integration approach and can be used with standard web service toolkits [17]. As the connectivity of OPC UA is based on this service-oriented approach, the security and reliability of the OPC UA communication is assured. The security concept is enriched by the AAA (Authentication, Authorization and Accounting) paradigm and used for every connection established between client and server independent of the technology mapping that has been chosen for data transport.

3.2 Using Artificial Intelligence to Learn from Data

The interoperability enabled by OPC UA-driven interface approaches leads to a resolution of the requirements needed for establishing a SOA-inspired architecture within industrial and automation environments without losing the deterministic behavior of field level systems. Regarded from a broader perspective, the arising system architecture brings together the ideas of loosely-coupled (service oriented) systems and tightly-coupled systems based on real-time capable deterministic behavior (as preferable at the shop floor level). The combination of these—on the first sight—contrary paradigms leads to a system behavior that is suitable for agile, learning and self-optimizing processes. These processes use process knowledge to adapt their behavior based on service-oriented communication while maintaining real-time behavior in its execution tasks.

The process knowledge that needs to be gathered to enable such learning complies with an overall information model that is, as defined above, ensured by the semantic communication and the interface standards. On the other hand, the usage of this knowledge base in-process requires flexible, service-oriented communication. The realization of SOA within automation environments is commonly performed by making use of multi-agent systems (MAS), in which agents serve as intelligent entities in the production environment. Similar to the SOA approach, every MAS provides their functionalities as remotely callable service [8]. This communication mechanism serves as a basis for scalable, distributed systems that are capable of organizing and solving complex tasks autonomously.

However, the practice of the communication in SOA-like MAS environments is often based on rather proprietary protocols or messaging conventions such as the Agent Communication Language (ACL) developed by the Foundation of Physical Agents (FIPA) [43]. Furthermore, the communication behavior is mostly based on a centralized orchestration of services or decentral organized choreography with rather fixed services.

In order to exploit the potential of such service-oriented, intelligent systems in its entirety, the communication between intelligent entities has to be combined with approaches for semantic communication build for machine-to-machine communication. This is done by lifting the agent communication on the next level—by making them talk “OPC UA”. This approach does not only formalize “intelligent communication” between smart entities in the form of well-established machine-to-machine communication, it also enables recombination of process knowledge that is gathered in low-level automation systems and that is generated by learning and intelligent entities on a meta-level of the production. The central advantage of this approach is that agent communication effectively operates on the same semantic level as arbitrary devices of the manufacturing environments that are using OPC UA.

4 Towards Interoperability in Human-Robot Interaction

Based on the performance benefits of autonomous and semi-autonomous technical systems over the past 10 years, the application range of technical systems in general and robotics in particular is greatly expanded. The documents “The German Standardization Roadmap” and “A Roadmap for US Robotics—From Internet to Robot” show how the penetration of professional and private life is dominated by robotic systems, which are based on a strong information network. Therefore, fundamental work is needed in the human-machine interaction. A robot must be developed that can be used in private and professional environment e.g. to support elderly to be further independent and productive and enable social participation.

In 20 years a robot similar as the mobile phone today will be our constant companion. To gain this objective there must be further development in the field of mechanics and energy, but most important for safe interaction is how the robot cooperates with humans and other machines. Therefore, the robot must recognize intentions of people to support them and the robotic system must make clear about its behavioral intentions to secure interaction. In this connection the main challenge is that human intentions consisting largely of non-verbal communication. For this purpose, an information model for robotics is needed that aggregates the atomic capabilities (rotational and translational motion) of the robot to more complex skills and associates those with context-sensitive strategies for functional performance. Furthermore, this mapping of capability and functional performance must be transparent and comprehensible for the human, so that in fact intentions can be derived from the procedures.

In order to achieve the outlined objectives a robot must have knowledge of its own functionality and the environment it interacts with. Therefore, on the one hand the system design has to be regarded, on the other hand capabilities as gathering, recognition and evaluation and respond need to be implemented accordingly. We take a closer look into these topics in the following sections.

4.1 System Design

In the system design the robot configuration as well as the security evaluation of robots and applications will be considered. Methods must be developed for safe human-machine interaction and the way of description of the interaction has to be determined. Based on these descriptions and configurations a secure robot configuration must be derived in order to detect reliable risks and unwanted interactions properly.

4.2 Gathering

For a safe human-robot interaction it is necessary that the machine can detect the human and the environment. The robot's detection of its environment is done by sensors and their data recording. Sensor values must be mutually correlated to provide comprehensive and verified information. The detection should be as good and efficient as possible, even under unfavorable conditions (for example lighting conditions, surfaces, fouling), to guarantee the safety of the human. Since a few years, optical sensor systems are also available to help monitoring the common working space of humans and robots in order to avoid collisions. However, this working area surveillance from a fixed perspective is the origin of general limitations such as masking the visible area. It can't be excluded that the worker is hidden behind a sensor. The robot can solve this fundamental problem by detecting obstacles in its range of motion by itself, so the "source of danger" is fitted with sensor technology. In this case spatial resolution and real-time capability of the sensor are special challenges for technical integration.

4.3 Recognition and Evaluation

Based on the sensor data a detection and an evaluation of the environment in which the robot and the human work is done. For this purpose, we need a reliable distinction of humans from other objects. Starting from the recognition of a scenario each situation must be evaluated with regard to whether this is an intentional interaction or a risk to humans. It must be sure that the algorithms that are used recognize and rate situations correctly. This project will implement a model that automatically transmits a movement shown by a person to an arm and a hand of an industrial robot. In the collaboration with robots, the people's safety is most important. Safety aspects of the generated robot programs are based on the individual workspace of various robot kinematics, the configuration of the robot motion commands avoid singularities and discontinuities. The human movements are divided into movement patterns. Motion analysis calculates all parameters of the robot commands in order to prevent dangerous situations. The robot learns movements and can apply these to any object. The movements are not copied but adapted to each task. The trajectory is adjusted in relation to the object location. The general object manipulation requires haptic understanding in order to quickly create a trajectory and handling movements for complex components in the required location.

4.4 Respond

A reaction of the robot is the response to a certain situation. How the robot reacts depends on the evaluation of the situation. The reaction always has to happen in a way that no danger to humans arises. The robot also has to react to misbehavior of the people so that there is no danger to him, or at least threads should be minimized. The reaction is related to the actuators which have to make sure that the control signals are error-free, i.e., when the expectations have been met, to start the reaction. During the movement, the path planning determines which reaction takes place in order to avoid collisions. In this case, however, the ability of the actuators must be known and considered.

Unlike robotic systems in production the direct interaction between humans and robots is an essential part when it comes to service robots. Especially while using mobile service robots, for example assistive robots in industrial or domestic environment, people must be safe while moving in their vicinity. Wherein the insertion of separated safety guard-devices is generally not possible. Mobile service robots with manipulators must frequent people's environments safely. The robot's workspace must be monitored three-dimensionally and its movements must be adapted to existing static and dynamic obstacles. The information model provides a method to generate an obstacle model out of the three-dimensional sensor data. This model is used for the trajectory planning of the manipulator and the mobile platform of the robot. The model includes all degrees of freedom of the robot system that includes the platform and the manipulator in order to enable simultaneous movements of manipulator and platform and avoid collision.

5 Use Cases

5.1 Adaptable Demonstrator for Flexible Production Organization

An increasing interest of leading manufacturing enterprises in experiencing developments of ‘Industrie 4.0’ is the elaboration of suitable demonstrators. These use-cases are of major importance to show the potentials of the fourth industrial revolution and its applications. Firstly, they serve as testbeds for the realization of visionary approaches, secondly they are used for the assessment of intelligent behavior in a manufacturing environment such as the deployment of MAS. One of these demonstrators has been carried out by several German universities, of whom the most important ones are the Technische Universität of Munich (TUM), the universities of Stuttgart and Augsburg as well as the RWTH Aachen University. The according prototype is referred to as “myJoghurt” [26]. The technical setup was carried out by and is located at the Institute of Automation and Information Systems (AIS) at the TUM [34].

The goal of this demonstrator is to show the potentials of ‘Industrie 4.0’ from the participating universities point of views while the different universities work out different parts of the demonstrator and enlighten different aspects of the production optimization of the process. The key element of the demonstrator consists of the usage of autonomous software agents to organize the production process independently of human intervention. The underlying MAS intents to organize (or accordingly reorganize) the production process in an autonomous way while intelligently dealing with errors or malfunctions [34]. The production flow is mapped from the start of production, which is initiated by accepting orders of possible customers, to the delivery of the final product [26]. During this process the customer, who initiates the process by placing an order, is able to regard the evolving of his product (the yoghurt), while the production machinery performs the process chain in an autonomous way.

In order to further improve the capabilities and far reaching reputation of the demonstrator, one target goal of further developments consisted in an enhancement of the communication process between the agents by enabling data exchange using OPC UA. The presented use case consists of OPC UA based software agents that communicate with all the other entities of the production network in a consistent way. At the RWTH Aachen University, a virtual software demonstrator has been carried out based on these approaches by enabling not only intelligent behavior of the agents, but also intelligent communication between these smart entities. Accordingly, an

OPC UA based representation of intelligent software agents has been developed to show the potentials of a "holistic" communication that is compliant to the data exchange in low-level production systems while remaining the service-oriented information exchange methods of the multi-agent system.

For the implementation of OPC UA services and message exchange methods that comply with the functioning of MAS, we had to take into account all described capabilities and abilities of an MAS. In terms of this work, we carried out an architecture for the embedding of multi-agent systems with OPC UA considering all system abilities of traditional MAS. Within this architecture, the agents of the embedded MAS, represented by OPC UA clients, connect to a central OPC UA server that functions as the Agent Management System (AMS) and also contains the abilities and storage functions of the Directory Facilitator (DF). The OPC UA server manages and coordinates a list of the registered agents with their abilities and at the same time organizes the production flow as well as the exchange and mapping of messages between the different agents.

By only routing the messages between agents, the OPC UA server sustains the communication autonomy of the intelligent software agents. The proposed architecture also takes into account a possible communication between multiple MAS by an inclusion of a mediation agent that organizes the information exchange between one AMS and main server instances of other MAS. The mediation agent guarantees a compliance of messages with the ACL standard. The exchange of messages between OPC UA clients and the central OPC UA server in terms of an automation environment is visualized in Fig. 3.

In the figure the information exchange between an intelligent agent and the OPC UA central server instance is shown in the context of a real production environment. The agent with the abilities of an OPC UA client connects to an OPC UA server on the field level, which propagates e.g. sensor data. After receiving data from the field, the agent accordingly creates messages, e.g. a write-message as shown. The content of this message including possible meta-data according to the underlying

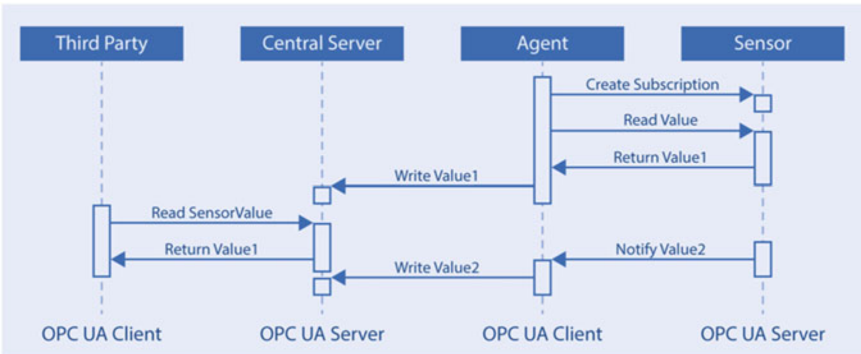


Fig. 3 Message transport flow chart between OPC UA agents representing machines on the shop floor and the central OPC UA server as agent management system

information model and additional system information can subsequently be further propagated by an agent by making use of the central OPC UA server as a mediator. The central OPC UA server can either coordinate the message to other agents or provide data access to other third party applications. These third party applications may be represented by enterprise wide information systems like ERP or MES. This ability of bringing together tightly-coupled systems from the shop floor and loosely-coupled service-oriented systems by making use of software agents as a link bears the real potential of the proposed approach as it finally tackles the true obstacles of interoperability in automation systems.

Based on the proposed OPC UA agent representation approach a generic demonstrator for production optimization has evolved that simulates the optimization of a production process by smart entities solely by communicating through OPC UA. The demonstrator maps models of machines on a network of distributed devices (Raspberry Pi), which represent the behavior of these machines in an intelligent and autonomous way (see Fig. 4).

The intelligent embedded devices (Raspberry Pi) shown as black boxes in Fig. 4 are capable of representing arbitrary machines and to organize themselves in an autonomous manner. The full demonstration process has been implemented for the above explained myJoghurt use-case. Based on an order of a customized yogurt by a customer the production process is initiated and organized by the smart entities. All information that is exchanged during the virtual production process is mapped on an OPC UA information model address space. The production process can be further influenced by the user through the interaction buttons shown at the bottom of the picture. While a production process is running, pushing the green button leads to an immediate finish of the current production step. Hereinafter the agent

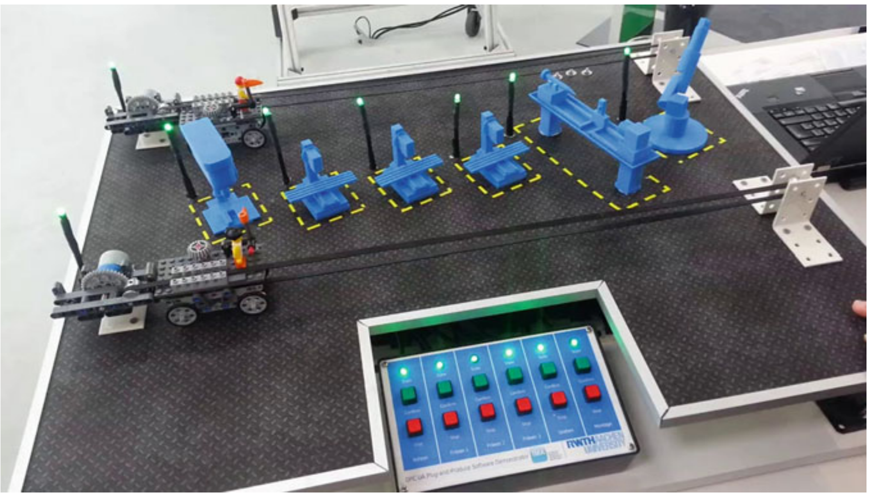


Fig. 4 Agent-based demonstrator simulating virtual production processes for intelligent agent-based manufacturing organization

network will reconfigure the production scheduling accordingly and continue the production process. The second interaction capability is initiated by pushing the red button, which is used to mark certain machines as “defective”. If a machine that is currently working on a production step is deactivated by pushing the red button, the network of intelligent agents will reorganize the production process considering machines that can replace the defective production machine to fulfill this particular production step.

The states of the production including work status of machines, transportation processes or production errors can be accessed through OPC UA clients. A web application was developed for the demonstrator to show these data access capabilities. Through the web app, the production can be configured and the production process is tracked. It shows a detailed overview of a currently running production steps, while some other steps are already shown as finished. Further production steps are awaiting the finishing of preliminary steps and are also visualized within the web application. Furthermore, the customer is able to see the specifications of the product currently manufactured and sees production times that remain until the customized product is finished. The web application relies on the namespace of the OPC UA server application and can be accordingly adapted to map any further use-cases by exchanging the namespace and address space model of the underlying OPC UA server.

5.2 *Virtual Production Intelligence Platform*

Another exemplary use case, which at the first glance, does not fit into the topic of interoperability in automation of cyber physical systems has been carried out in the domain of virtual production. Virtual production or the computer-integrated manufacturing (CIM) as superordinate term, refers to multiple functional areas as design, analysis and planning as far as factory floor functions such as providing direct control and monitoring of all operations. One important and, not successfully implemented show-stopping aspect regarding CIM, is the uniting of the data processing and the sensor components using suitable methods. At least this is one of the reason why CIM failed in the 1970s. In this use case, we addressed this emerging issue with Adaptive Information Integration [30].

The virtual production intelligence (VPI) platform addresses the interoperability problems arising when unifying components and applications along different domains. Thereby, it does not address propagation respectively communication issues, but moreover the problem of heterogeneity. Heterogeneous, distributed applications, sensors and other components use different descriptive formats and structures to describe digital models of various processes. Consolidating and transforming data automatically into the requested format or structure by taking the semantics into account is the goal of the VPI platform. In our work we successfully applied the method of Adaptive Information Integration to different domains, respectively factory planning [6], laser cutting processes [36], analysis of material

Fig. 5 Screenshot of the VPI platform with selected factory planning domain (KPI cockpit)

behavior [28] as well as consolidation of machine data [18] and consolidated it within the VPI platform (cf. Fig. 5).

VPI denotes our concept that enables product-, factory-, and machine planners to plan products and their production collaboratively and holistically [37]. The concept comprises methods to consolidate data generated in the aforementioned domains. It also includes visualization and interaction techniques to analyze and to explore the retrieved data. As previously mentioned, heterogeneity still represents one of the main challenges in information integration. Thereby, the VPI concept does not pursue standardization. Instead, this approach formulated new methods based on the aforementioned Adaptive Information Integration. The concept in further detail is presented in Meisen [28].

The main objective of the approach is to remove the need for specialized adaptors to bridge the gap between the data model of the source and the one of the sink. We use a domain-specific information model to map the underlying conceptual bases and combine this process with a service bus architecture. We use different ontology technologies to transfer the parts of the information model that get lost during the translation into the logical data model to a domain-specific ontology. Further, we define mappings between the data model and the concepts of the ontology [30]. In addition, domain-related constraints and axioms that are also modelled along the concepts are used to validate data handled by the system and to enrich the presented data with implicit information.

The aforementioned procedure model along with the implemented semantic technology enables the information system to understand data in sense of semantical concepts. Hence, the system can automatically retrieve further implicit information, check the data for validity and consolidate it respectively. Furthermore, the data can automatically be translated in other structures or formats, just by defining and adding further capabilities to the system. The required transformation process to close the gap between the available data and the requested data can automatically be determined by the system using the description of capabilities and the semantically annotated data.

5.3 Canoe

One other aspect of interoperability in the factory of the future is to maintain the interaction capabilities of human workers with the machinery. Thus, a skilled worker should still be able to use robots optimally, therefore the robot must have the capability of self-description. It must aggregate its "atomic" operations to context-sensitive and understandable actions, thus the process of movement from one to another discrete point and the opening and closing of a gripper means "Handling an object". Furthermore, should the robot system be able to transfer knowledge from one context to another context, so that a system does not stick to the actual action, but on the function that has to be fulfilled. For example, a robot system performs a surface treatment by moving a nozzle along a trajectory while a coating is applied to a metal sheet. In the classical approach, the robot and the tool (nozzle) can be programmed according to the task. Now, however, the robot has knowledge about which tool it can use for any purpose that need to be integrated with the specific knowledge of the robot programmer. A Toll can be used for a specific manipulation of the environment. Tool and manipulation will be derived from the desired target state, then a procedure and therefore programs are being generated. The actions of the robot are no longer based on the starting points, but on the knowledge, i.e. the ability of "painting". To achieve this goal, the robot must have knowledge about the functionality on its own and the environment, as well as the interaction with it.

These developments make it possible to the transfer the "classical industrial robotics" to human-centered assistance systems that can be used both in domestic and occupational environments. Since large-scale industries have already achieved a high degree of automation, the application scenario is taken out of the group of SMEs. In the described scenario, the production of canoes will be automated. The recreational boats are manufactured in a plastics processing process of PE. The result of this proves is a canoe-blank that needs to be post-processed in further steps. These post-processing steps are surveying the boat, milling out of surfaces, drilling holes and deburring of milling and drilling edges. These steps are currently performed manually and are subject to the time and quality variations of manual processing. The aim is to automate these steps with a robot-based man-machine interaction system. Figure 6 shows a canoe and associates it with the post-processing challenges that have so far prevented automation. The PE component is subject to manufacturing conditional shrinkages, in a way that each component in its extension is different from other boats and CAD data. Therefore, each boat has to be detected

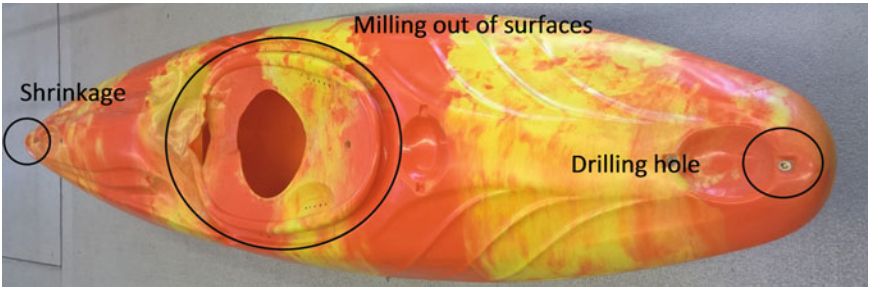


Fig. 6 Canoe manufacturing

with its unique dimensions that are not known a priori. The partly matt and partly reflecting surfaces as well as unstructured gradients further complicate an optical detection of the component. However, the milling, drilling and deburring are performed depending on the geometric data and position in space.

A human-machine interaction system will be set up that fulfill these tasks in collaboration with human workers. To achieve a high degree of process reliability, timing and consistent quality, the handling of the boat, a scanning of the boat as well as processing steps will be done by an implementation of a convertible robot-based production cell. In order to handle unknown system conditions that may occur during the manufacturing process the robot system will be monitored and coordinated by workers. The production of this PE-component was selected for the scenario, because the handling of the above mentioned reasons is very demanding.

6 Conclusion

Interoperability is the key enabler for the realization of smart automation. The approaches that were shown—based on machine-to-machine interaction and human-machine interaction—deliver key components to make yet dedicated systems and human beings in the area of production grow together. These extended interactions and communication capabilities enable the recombination of system configurations from different areas of the production having different requirements in terms of real-time functionality, robust behavior and flexible organization. The demonstrated approaches are suitable of linking tightly-coupled systems, which are necessary on the shop floor level, with loosely-coupled systems for a flexible service-oriented communication and interaction of highly distributed systems. In the described use cases we have shown the successful application of our methods and procedure models. In the future, we will extend our pool of demonstrators especially in other domains.

Regarding OPC UA we will focus on an advancement of modeling approach for MAS (e.g. enhancement of autonomous behavior) and on a development of the ‘Industrie 4.0’ demonstrator as a suitable proof of concept for the derived approaches. In terms of the scientific research on MAS, the agent-based approach will be advanced, especially in terms of far reaching autonomy, self-organization, self-adaptation and learning of the agents. The integrated communication through OPC UA intends to serve as an interface to match the needs of the different agent applications with high flexibility and adaptability. The centralized coordination approach will be replaced by completely independently-acting agents which work in accordance to the black board pattern. In terms of the modeling in OPC UA, the next-generation agent will use generic OPC UA methods to subscribe to production channels. For this purpose, e.g. the Alarms & Events specifications can be employed for real-time process and machinery diagnostics.

In case of any unforeseen events during the production, the agents will further autonomously react on new situations and offer their abilities to solve the detected issues. This proactive behavior results in an artificial intelligence that is learned throughout the production process. We will further extend this behavior up to the human interface, so that a real human-machine interaction will be within reach in the next years.

References

1. Ajzen I (1991) The theory of planned behavior. Organ Behav Hum Decis Process 50(2):179–211
2. Bagozzi R, Baumgartner J, Yi Y (1989) An investigation into the role of intentions as mediators of the attitude-behavior relationship. J Econ Psychol 10(1):35–62
3. Belkin NJ, Oddy RN, Brooks HM (1982) Ask for informational retrieval: Part I. background and theory. J. Doc. 38(2):61–71. doi:10.1108/eb026722
4. Brown N, Kindel C (1998) Distributed component object model protocol—DCOM/1.0. Microsoft Corp
5. Budden JS, Sagarin BJ (2007) Implementation intentions, occupational stress, and the exercise intention-behavior relationship. J Occup Health Psychol 12(4):391–401
6. Büscher C (2015) Ontologiebasierte Kennzahlentwicklung für Virtual Production intelligence, Als Manuskript gedruckt. Schriften zur Informations- und Kommunikationstechnik, vol 846. VDI Verlag, Düsseldorf
7. Chatzisarantis NLD, Hagger MS (2007) Mindfulness and the Intention-behavior relationship within the theory of planned behavior. Pers Soc Psychol Bull 33(5):663–676
8. Dijkman R, Dumas M (2004) Service-oriented design: a multi-viewpoint approach. Int J Coop Info Syst 13(04):337–368. doi:10.1142/S0218843004001012
9. Dragan A, Gordon G, Srinivasa S (2011) Learning from experience in manipulation planning: setting the right goals. In: Proceedings of the international symposium on robotics research (ISRR)
10. Dragan A, Holladay R, Srinivasa S (2014) An analysis of deceptive robot motion. In: Robotics: science and systems, p 10
11. Dragan A, Srinivasa S (2012) Online customization of teleoperation interfaces. In: 21st IEEE international symposium on robot and human interactive communication (Ro-Man), pp 919–924

12. Enste U, Mahnke W (2011) OPC Unified Architecture: Die nächste Stufe der Interoperabilität. at - Automatisierungstechnik 59(7):397–404
13. Ewert D, Mayer MP, Schilberg D, Jeschke S (2012) Adaptive assembly planning for a nondeterministic domain. In: 2012 AHFE international conference, pp 2720–2729
14. Ewert D, Schilberg D, Jeschke S (2012) Selfoptimized assembly planning for a ROS based robot cell. In: Hutchison D, Kanade T, Kittler J, Kleinberg JM, Mattern F, Mitchell JC, Naor M, Nierstrasz O, Pandu Rangan C, Steffen B, Sudan M, Terzopoulos D, Tygar D, Vardi MY, Weikum G, Su C, Rakheja S, Liu H (eds) Intelligent robotics and applications, vol 7506. Springer, Berlin, pp 696–705
15. Fishbein M, Ajzen I (1975) Belief, attitude, intention, and behavior: an introduction to theory and research. Addison-Wesley, Reading
16. Gruber TR (1993) A translation approach to portable ontology specifications. Knowl Acquis 5 (2):199–220
17. Hannelius T, Salmenpera M, Kuikka S (2008) Roadmap to adopting OPC UA. In: 6th IEEE international conference on industrial informatics, pp 756–761. doi:10.1109/INDIN.2008.4618203
18. Hoffmann M, Büscher C, Meisen T, Jeschke S (2016) Continuous integration of field level production data into top-level information systems using the OPC interface standard. Procedia CIRP 41:496–501. doi:10.1016/j.procir.2015.12.059
19. IEC (2015) OPC unified architecture (62541). https://opcfoundation.org/developer-tools/specifications-unified-architecture. Accessed 25 Sept 2015
20. IEEE (1990) IEEE standard glossary of software engineering terminology
21. Jasperneite J, Feld J (2005) PROFINET: an integration platform for heterogeneous industrial communication systems. In: 2005 IEEE conference on emerging technologies and factory automation (ETFA 2005), pp 815–822
22. Lange J, Iwanitz F, Burke TJ (2010) OPC: from data access to unified architecture, 4 rev edn. VDE-Verl, Berlin, Offenbach
23. Leitner S, Mahnke W (2006) OPC UA—service-oriented architecture for industrial applications. ABB Corporate Research Center
24. Mahnke W, Damm M, Leitner S (2009) OPC unified architecture. Springer, Berlin [u.a.]
25. Manning CD, Raghavan P, Schütze H (2009) Introduction to information retrieval. Cambridge University Press, New York
26. Mayer F, Pantförder D, Diedrich C, Vogel-Heuser B (2013) Deutschlandweiter I4.0-Demonstrator: Technisches Konzept und Implementierung
27. Mayer MP, Odenthal B, Ewert D, Kempf T, Behnen D, Büscher C, Kuz S, Müller S, Hauck E, Kausch B, Schilberg D, Herfs W, Schlick CM, Jeschke S, Brecher C (2012) Self-optimising assembly systems based on cognitive technologies. In: Brecher C (ed) Integrative production technology for high-wage countries. Springer, Berlin, pp 894–946
28. Meisen T (2012) Framework zur Kopplung numerischer Simulationen für die Fertigung von Stahlerzeugnissen, Als Ms. gedr. Fortschritt-Berichte VDI: Reihe 10, Informatik, Kommunikation, vol 823. VDI-Verl., Düsseldorf
29. Meisen T, Meisen P, Schilberg D, Jeschke S (2011) Application integration of simulation tools considering domain specific knowledge. In: Proceedings of the 13th international conference on enterprise information systems. SciTePress, pp 42–53
30. Meisen T, Meisen P, Schilberg D, Jeschke S (2012) Adaptive information integration: bridging the semantic gap between numerical simulations. In: Zhang R, Zhang J, Zhang Z, Filipe J, Cordeiro J (eds). Enterprise information systems: 13th international conference, ICEIS 2011, Beijing, China, 8–11 June 2011, Revised Selected Papers. Springer, Berlin, pp 51–65
31. Meisen T, Reinhard R, Schilberg D, Jeschke S (2011) A framework for adaptive data integration in digital production. In: Spath D, Ilg R, Krause T (eds) Innovation in product and production: conference proceedings, July 31–August 4 2011 in Stuttgart, Germany. Fraunhofer-Verl., Stuttgart, pp 1053–1066

32. Meisen T, Rix M, Hoffmann M, Schilberg D, Jeschke S (2013) A framework for semantic integration and analysis of measurement data in modern industrial machinery. In: Proceedings of the international symposium on measurement technology and intelligent instruments (ISMTII 2013). Apprimus-Verlag, Aachen
33. Object Management Group (2015) Data distribution service (DDS). http://www.omg.org/spec/DDS/1.4. Accessed 27 Oct 2015
34. Pantförder D, Mayer F, Diedrich C, Göhner P, Weyrich M, Vogel-Heuser B (2014) Agentenbasierte dynamische Rekonfiguration von vernetzten intelligenten Produktionsanlagen – Evolution statt Revolution. In: Bauernhansl T, Hompel M ten, Vogel-Heuser B (eds) Industrie 4.0 in Produktion, Automatisierung und Logistik. Springer Fachmedien Wiesbaden, Wiesbaden, pp 145–158
35. Pardo-Catellote G, Farabaugh B, Warren R (2005) An introduction to DDS and data-centric communications
36. Reinhard R (2015) Visual Analytics für die Virtuelle Produktion. Logos Berlin, Berlin
37. Reinhard R, Büscher C, Meisen T, Schilberg D, Jeschke S (2012) Virtual production intelligence—a contribution to the digital factory. In: Su C, Rakheja S, Liu H (eds) Intelligent robotics and applications: 5th international conference, ICIRA 2012, Montreal, QC, Canada, 3–5 Oct 2012, Proceedings, Part I. Springer, Berlin, pp 706–715
38. Rinaldi J (2013) OPC UA—the basics: an OPC UA overview for those who are not networking gurus. Amazon, Great Britain
39. Tovar E, Vasques F (1999) Real-time fieldbus communications using Profibus networks. IEEE Trans Ind Electron 46(6):1241–1251. doi:10.1109/41.808018
40. Twinoaks Computing Inc. (2011) What can DDS do for you?: Learn how dynamic publish-subscribe messaging can improve the flexibility and scalability of your applications
41. Vieritz H, Yazdi F, Göhner P, Schilberg D, Jeschke S (2011) User-centered design of accessible web and automation systems. In: Holzinger A, Simonic K (eds) Information quality in e-Health: 7th conference of the workgroup human-computer interaction and usability engineering of the austrian computer society, USAB 2011, Graz, Austria, 25–26 Nov 2011, Proceedings. Springer, Berlin, pp 367–378
42. W3C OWL Working Group (2009) OWL 2 web ontology language
43. Wooldridge MJ (2009) An introduction to multiagent systems, 2nd edn. Wiley, Chichester

Enhancing Resiliency in Production Facilities Through Cyber Physical Systems

Robert Schmitt, Eike Permin, Johannes Kerkhoff, Martin Plutz and Markus Große Böckmann

1 Introduction

Cyber-Physical Production Systems (CPPS) as a derivate of a Cyber-Physical System (CPS) in a manufacturing environment represent the next stage of development of production technology. The dissolution of hierarchical system structures induced by this technological trend will lead to reconfigurable networks of Smart Products and CPPS in the future. The resulting complexity creates a new demand for manufacturing companies: robustness of their production systems with a high adaptability at the same time. These two aspects are the core characteristics of the resilient factory.

In contrast to flexible production systems, which provide a stable area of ability and are only scalable within fixed limitations, a versatile production system provides changes in case of need and possesses possibilities to rebuild and extend functions as an essential basic property [32]. Nowadays, flexible production systems can be realized through the use of automation solutions and organizational

R. Schmitt (✉)
Laboratory for Machine Tools and Production Engineering (WZL),
RWTH Aachen University, Steinbachstraße 19, 52074 Aachen, Germany
e-mail: r.schmitt@wzl.rwth-aachen.de

E. Permin · J. Kerkhoff · M. Plutz · M.G. Böckmann
Fraunhofer Institute for Production Technology (IPT),
Steinbachstraße 17, 52074 Aachen, Germany
e-mail: eike.permin@ipt.fraunhofer.de

J. Kerkhoff
e-mail: johannes.kerkhoff@ipt.fraunhofer.de

M. Plutz
e-mail: martin.plutz@ipt.fraunhofer.de

M.G. Böckmann
e-mail: markus.grosse.boeckmann@ipt.fraunhofer.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_11

concepts such as Lean Production. Cyber-Physical Production Systems provide the opportunity to allow the adaptability in the sense defined above and thereby to open up productivity potential in companies. The adaptability thus causes the ability of using the added value for the inactive percentage of potential productivity through rebuildable functionalities elsewhere.

Another important aspect is the Plug and Produce-capability of Cyber-Physical production systems, which is made supported by an autonomous self-monitoring. However, the required decentralized networking of CPPS and Smart Products involves risks for the robustness of the production. How these safety and security risks are to be handled remain open questions and are addressed in this contribution.

2 The Need for Resilient Factories in the Context of Industry 4.0

The following subsections illustrate why resilient plants are needed in the Industry 4.0, what is meant by this term and which scientific foundations play a role.

2.1 Cyber-Physical Systems, Self-Optimization and the Internet of Things

2.1.1 Cyber-Physical Systems (CPS)

The term Cyber-Physical Systems was originally coined by Helen Gill of the National Science Foundation in 2006 [28]. Cyber-Physical Systems describe the networking of ICT-systems among each other and with the internet [12]. A general definition was provided by Lee, who referred to CPS as the integration of computer calculations in physical processes. Physical processes and calculations mutually influence each other in a Cyber-Physical System by feedback [28].

Broy used a layer model to describe Cyber-Physical systems [8]. He integrated the "system in system"-concept as well, a hierarchical structure of systems, which in turn are an element of a superior system. Furthermore, one can conclude from these assumptions that embedded systems are not to be used synonymously with CPS. Cyber-Physical Systems rather constitute a framework around the embedded systems and allow a networking outwards at the same time.

In the *Research Agenda CPS* presented by the German Academy of Science *acatech* in 2012, Cyber-Physical Systems were described as the connection between physical and information technology world. CPS arise as a result of the networking and integration of embedded systems, infrastructure and application systems. The latter especially require human-machine interaction [9]. In CPS, a particular focus lies on sensors and actuators. While sensors are recording data from

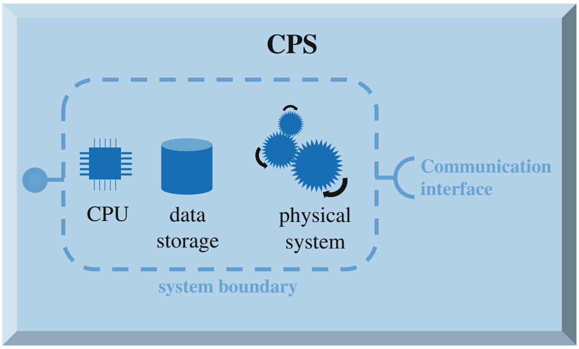


Fig. 1 Basic structure of a Cyber Physical System

the environment, actuators are ensuring the implementation of physical processes. Further characterizing features of Cyber-Physical Systems are the so-called self-X capabilities (e.g. self-description), an increased adaptability as well as a significantly increased local intelligence [4].

The simple illustration of a CPS is given in Fig. 1. This representation is taken from the project CyProS—productivity and flexibility enhancement through the networking of intelligent systems in the factory. An essential observation here is the combination of a physical system with a data storage and data processing function, which has extensive interfaces [17].

In the context of this article, both Cyber-Physical Transport Systems (CPTS) and Cyber-Physical Assistance Systems (CPAS) are relevant. This refers to CPS which is only responsible for the Transport of Goods (Transport) as well as the Support of Employees (Assistance). CPTS and CPAS alone do not change or affect raw materials.

2.1.2 Cyber-Physical Production System (CPPS)

A specialization of CPS in a manufacturing context are Cyber-Physical Production Systems (CPPS). These are CPS which are integrated in the industrial environment, also in particular the product, the production and the Production Systems (CPPS) [25]. CPPS should help businesses create a self-organizing adaptable production by simplifying change processes. Standardization and modularization play an essential role here because they allow a company-wide connection.

A CPPS consists of at least one CPS, a production planning software (PPSW) and can also contain CPTS and CPAS (see Fig. 2). Contrary to the technical status quo, the PPSW does not take over a complete controlling and planning function for the whole production. Instead, it rather provides a superior objective function, the basis of the entities of the system. Beyond that, the CPPS can exist out of more CPPS in terms of a hierarchical structure with sub-systems.

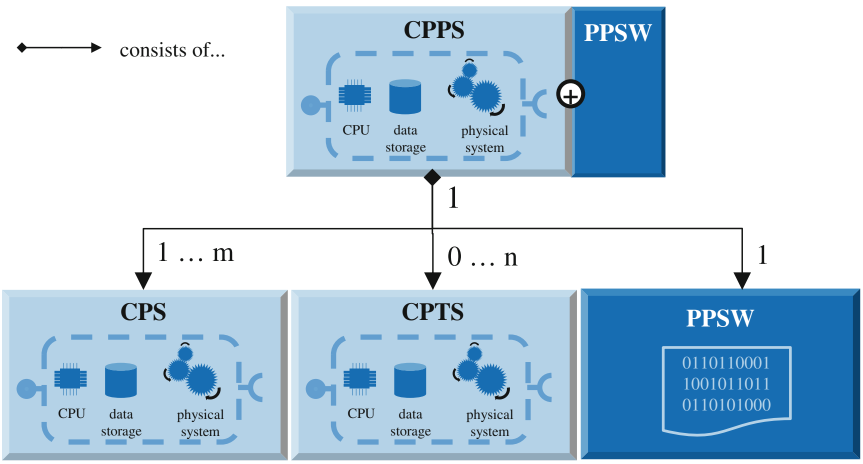


Fig. 2 Cyber Physical Production System

2.1.3 Self-Optimization

The term of Self-Optimization has been characterized through the collaborative research center 614 "Self-Optimizing Systems of the mechanical engineering" in Paderborn. According to Gausemeier and Adelt, the Self-Optimization is accomplished in three steps: The analysis of the status quo, the determination of the objectives and concluding the adaptation of the system performance [1]. The optimization itself is executed alternatively in the second step to examine relevant goals or in the third step to realize favorable adaptions [22].

In the course of the Cluster of Excellence "Integrative production engineering for high-wage countries" in Aachen, the application of such systems in manufacturing has been examined and achieved [6]. Self-Optimization can be distinguished from classic control circuits essentially by their ability of the adaption of the objectives of the systems, the controller as well as the controlled system [36].

2.1.4 Internet of Things

The term Internet of things generally describes the coalescence of the virtual and the real worlds [12]. Bullinger and ten Hompel considered the Internet of things as a vision where part of the physical world and the real objects can be extended with additional information thanks to the Internet [11]. Due to the connection of machines, sensors and actuators to the Internet, they are developing to become protagonists themselves. This depicts a main feature of the Internet of things and also the Internet of services [25].

The Internet of services is to be regarded as an extension of the Internet of things. It complements the Internet with a service platform, which facilitates the development of web-based services [25]. At the moment, an increasing number of so-called Business Apps are being recorded. By using these, customers can combine different software components to highly flexible solutions. Especially Cloud Computing has to be mentioned in this context, as it offers computing capacity and storage/memory via the Internet as a service. Instead of buying and maintaining their own expensive server computers, companies can outsource the majority of their IT thanks to Cloud Computing [12].

2.1.5 Industry 4.0

From a present-day perspective, the first industrial revolution represents the introduction of mechanical production systems operated by external power sources such as the power loom in the 18th century. The second industrial revolution is being denoted as the introduction of mass production based on the division of labor and devices like band conveyors in the late 19th and beginning 20th century. With the application and use of electronics and automation of manufacturing plants starting from the 1970s, the third industrial revolution has finally been introduced [12]. In the fourth industrial revolution, the focus is now put on stronger connections of information and communication technology with the industry. Referring to the Web 2.0 the term Industry 4.0 is frequently being used.

In the course of this article, Industry 4.0-components are considered as Cyber-Physical System as discussed above, but with the following, additional characteristics: performance according to the Industry 4.0 semantics, communication and functionality based on the SOA-principle (service-oriented architecture), worldwide distinct identification (e.g. IP address), security functions, virtual self-description including dynamic performance as well as the opportunity of a permanent status update [25].

2.2 Resilient Production Systems

2.2.1 Market Trend Customized Products

To produce the products demanded by the customer, machines and facilities including their components and tools, storage and transportation means, the raw material, semi-finished products and construction parts as well as the employees need to be regarded as essential elements of the production systems in particular [47]. Customer demands are subject to continuous changes and in the last years developed to form a buyer market driven by demands, not offers.

Thus, customized products for manufacturing companies of highly industrialized states such as in Germany are becoming an important enabler to exist in the face of global competition. The inherent complexity of these products places new demands to those companies and their factory: The adaptability of their production systems as the most important requirement, to efficiently produce smallest batch sizes down to one-piece-flows.

This development of customer individual production will further be driven by the future horizontal integration over value creation networks through a reconfigurable ad-hoc connection of producers with their customers and suppliers in context of the Industry 4.0.

Previous technical approaches to tackle these demands have arisen from the field of automation on the one hand and are supported by organizing concepts such as Lean Manufacturing on the other hand. Still, the combination of these existing technologies will not be able to cope with the future demands towards flexibility and adaptability:

Both solution approaches have in common that the flexibility of production systems is being accepted as required technological basis for the achievement of a customized production. That means that the production systems are being developed in size and basic design to be able to cope with the entire range of requirements from the beginning. These again are determined by customer requirements for product specification, time of delivery and costs. Here, set-up processes to shift the production from one product variant to another is being optimized through automation solution and organizing concepts such as Single Minute Exchange of Die (SMED), to obtain the highest possible productivity of flexible production systems.

In the future the customer demand for individuality of products will also be enforced by horizontal networking via supply chains. It is conceivable that e.g. the excess production capacities in the Internet of things will be offered and will be used according to their available capacity by customers on appropriated Internet platforms. This leads to dynamic supply chains and—with existing opportunities of configuration and individualization through the customer—to continuing increasing product and production complexity.

2.2.2 Technology Trend Cyber-Physical Production Systems

To enable Cyber-Physical Production Systems to communicate and operate with each other, there is a need for models of the physical systems (machines and facilities as well as products), which contain a different description of the reality depending on their purpose: In development and design, models of the construction parts to be produced later mostly focus on geometrical, substantial and functional aspects. In production, other models of the product to be manufactured come to the foreground such as method description, operating instruction or NC programs. From a sales perspective, the quantity and cost models play a superior role. Models of the availability of machines and facilities need to be added, which can be utilized for the production planning or the simulation of production process. That means there are not only various product models of a company, but also various models of its resources (employees, machines, facilities, technical building services etc.), which are required for the manufacturing of the products.

This diversity of specific product and production system models as well as their interactions lead to an indefinite information situation. A main risk is that inconsistencies in several model representations lead to defects in the product development process. Furthermore, there are certain security risks to the cyber world depicted by the fact that the above-mentioned models that could be modified specifically due to taking advantages of security gaps in the Industry 4.0. As a result, Cyber-Physical Production Systems have to be designed in a robust way to be able to deal with inconsistencies in the model world.

2.2.3 Global Trend of Energy and Resource Efficiency

The worldwide persistent population growth and the scarcity of natural resources impose already today and in the future reinforced high requirements on the use of energy and resources. This applies to the private households, the public sector and in particular also the industry. Thanks to energy- and resource-efficient processes in the industrial production it is possible to save expenses and to increase the productivity.

2.2.4 The Resilient Factory as a Response to the Given Trends

Resiliency is an often applied concept in many scientific areas, which was shaped in the field of psychology and later on modified by biology, system theory and sociology [29]. In general, resiliency is the ability of a system to recover from an external perturbation and return to its initial state (see [41] or [23]). In organization theory, the term of resiliency is already in widespread use. In this field, resiliency is the immanent ability of an organization to maintain a stable state in face of a continuous external perturbation and/or occurrence of unpredicted major events like natural disasters [23]. This abstract definition could be easily scaled from the system boundaries of an organization, like a producing company, to the functional area of the production, but it receives further specification in the field of production engineering. Here, resiliency of production systems—especially in connection with Industry 4.0—is defined attributively with terms like persistence, adaptability, agility, redundancy, learning capability and decentralization [9]. Flexibility is an ability that is an implied resiliency characteristic [14]. Adaptability exceeds the concept of flexibility, it considers explicitly the functionalities to dismantle or expand, while flexible production systems solely provide a certain range of tolerance.

The market trend of mass customization and the technology trend of cyber physical production systems result in the adaptability and robustness as necessary core characteristics of future production facilities. Additionally, energy and resource efficiency form substantial characteristics of the Resilient Factory. This reflects that adaptability and robustness should not be enforced with undergoing excessive use of resources. Through those four characteristics we define the Resilient Factory within the framework of Industry 4.0 (see Fig. 3).

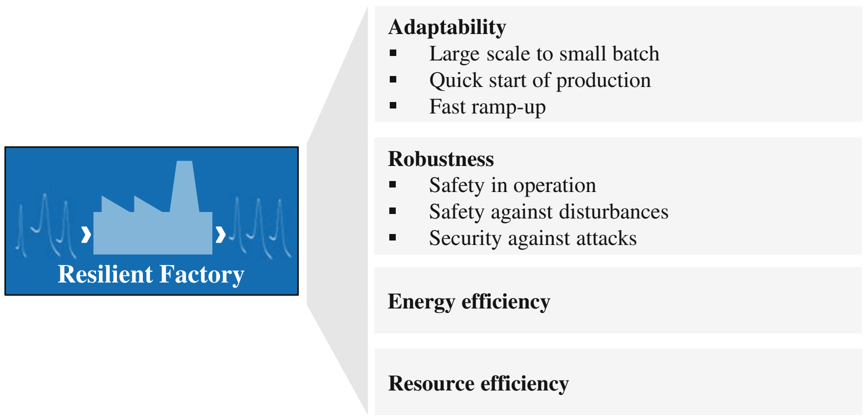


Fig. 3 Characteristics and abilities of a resilient factory

3 Objectives of the Resilient Factory

Through the characteristics defined in previous chapter, the Resilient Factory enables an increase of productivity in comparison to the current state of the art, which is exemplarily shown in Fig. 4. Adaptability, robustness as well as energy and resource efficiency represent the substantial requirements to better exploit existing resources and thus reduce waste.

In today's production facilities, productivity is wasted in many ways: unplanned perturbations of machinery and equipment result in delivery delays and waste. Planned downtimes of machinery and equipment, e.g. for setup processes of flexible

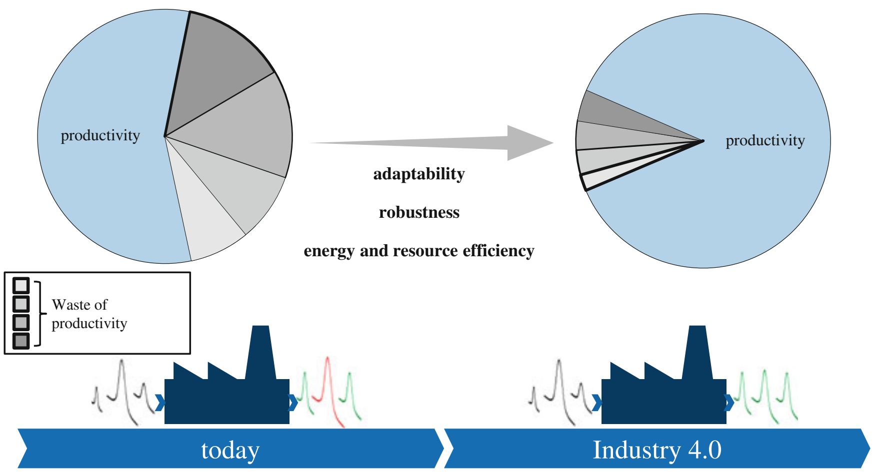


Fig. 4 Increasing the productivity as main goal of a resilient factory

production systems, are non-productive times. In addition to that, there are productivity losses according to the Toyota Production System: Mura (Japanese: unevenness) are losses which are caused by a disharmonious capacity utilization within the context of production control. Muri (Japanese: overburden) describes losses within work processes. On the one hand it refers to employees who produce errors due to physical or psychological overload and on the other hand it refers to machinery with long waiting lines due to inaccurate production planning [10, 26].

The most significant source of losses is waste (Japanese: muda), which divides into the following eight kinds [10]:

- overproduction
- waiting and idle time
- unnecessary or wrong processing steps
- unnecessary and long transportation ways
- large inventory
- unnecessary movements
- defects and defect consequences
- unused creativity of employees.

Additional waste in flexible production systems occurs through unused machine capabilities which are provided in accordance with the spectrum of requirements in design of the production system but not required from the product view. This range is defined by customer requirements and implies various product types and variants, which need to be manufactures within a certain time, with required quality and at adequate costs. Based on this the performance range of a production system is defined which includes certain ranges of capabilities of machines, storage and transportation systems as well as employee skills and expertise.

One goal of present production planning and control is the maximization of the capacities of the production systems. The applied indices usually set time factors in proportion to each other [46]. However, with regard to the available range of capabilities, this kind of optimization generates waste which is due to the fact that active abilities remain unexploited although those could be used regardless of the current processed order. Adaptability, as one of the two core characteristics of the Resilient Factory allows eliminating this waste and therefore leveraging the referred productivity potential. The adaptable elements of a production system need to be highly robust to be able to function consistently reliable in numerous contexts.

The previous implementation and the parameters introduced in Fig. 4 especially define the flexibility of production systems. But in the context of security concerns, productivity reserves exist, too. So, typically preparation steps are included in the setup time which ensures the security in the human machine interface. As those processes cannot be externalized in terms of setup time optimization, e.g. realization in parallel to the value-adding processes of the production system, the machine is not productive at this time [42]. This kind of security reserves will continue to exist in a Cyber-Physical production system as well as the human—against some current scientific and industry scenarios—will stay an essential part in the future production [43]. The occurring autonomous interactions within the production facilities or, in context of a vertical interconnection of production systems, through value chains in the internet of things and services arise in new challenges for the planning of the security measures, as introduced above. In this context, especially emergent patterns of self-organizing and self-optimizing technical systems play an important role [40] due to the consequences of not completely predictable interactions of Cyber-Physical Subsystems (compare chapter on scientific requirements for CPPS).

4 Two Example Cases for Cyber-Physical Systems in Production

In the following chapter, two example cases for Cyber-Physical Systems in production environments are presented. An energy-oriented manufacturing planning and control system as an example of a complete Cyber-Physical Production System will be discussed in the first section. It consists of sub-systems for production, transport and planning and thus illustrates cooperation and composition of CPPS. Afterwards, the potentials and benefits of smart glasses as Cyber-Physical Assistance Systems in industrial assembly tasks will be presented.

4.1 Energy-Oriented Manufacturing Planning and Control System

4.1.1 Requirements and Conditions

In the course of the fourth industrial revolution framework conditions of production process become more and more dynamic. Apart from customized products, decreasing stocks and resource efficient production technologies, future production systems have to deal with short terms changes in planning, both introduced customer-, supplier- and development-side. Moreover, production resources like energy that are characterized by their increasing prices as well as partly limited accessibility have to be managed actively in production planning of the future. Conventional production systems are not able to cope with such demands. In the following, decentral structured CPS are introduced as a suitable solution.

From a production planning point of view, Cyber-Physical Systems can be characterized as a combination of automatized processes that are set up by autonomous control loops and decentralized decisions. Orders and their starting times on machines are planned manually by an employee in a central control center. In the future, machines and technical systems will organize and autonomously handle their working content. This will cause a radical reduction of reaction times of control loops. For example in case of machine errors or order related delays, a re-planning will be triggered automatically. Nevertheless, the introduction of such production systems will not happen in a disruptive way, but gradually in most industrial sectors. Even in small and medium sized manufacturing enterprises the transformation will occur stepwise.

There are some parallels to the vision of autonomous car driving. This target seems to be achievable via intermediate stages of driver assistance technologies. Additional features like lane-departure assistant, active cruise control and traffic jam assistant, that are available even in current cars, ensure a stepwise, safe and based-on-experience development of the following technologies to reach the long term goal.

In detailed production planning and scheduling manufacturing execution systems (MES) are state of the art nowadays. In the conventional "Automation Pyramid" as well as in other concepts like the "Automation Diabolo" [44], which are conversely discussed in the course of Industry 4.0, MES is the connecting string between rough planning and enterprise resource planning (ERP) respectively product lifecycle management (PLM) level and MDC (machine data collection)/PDA (production data acquisition) systems at the shop floor. Generally, MES transforms middle and long-term capacity scheduling in daily production orders for certain machines under consideration of restrictions like personnel, machine utilization and tool availability. The manufacturing process is split up into several operations, each corresponding to one single working step. The majority of companies are using MES for simple process monitoring or manual planning of machine use, where planning paradigms are determined by experience and tacit knowledge of the employees. A rule-based production planning that is related to more than one target figure and its interdependences exist in the fewest companies. Unfortunately, to realize CPS in a production environment precisely this is necessary, so the currently existing obstacles will be discussed.

The first main challenge for decentralized decision making processes is a *continuous, real-time collection and provision of information*. It has to be secured that all necessary data to take well-grounded decisions are available both globally at the control station as well as locally at the machines. For that, standardized interfaces that are aligned vertically through the automatization pyramid, an adequate data granularity, memory availability and easy solutions to integrate machine-independent information like status reports into the existing data stream are needed.

Because of the exponentially growing amount of information and interdependences between machines, orders and operation materials, that can no longer cognitively processed by a single person, the second condition is the *development of high performance, multidimensional optimization algorithms*. Those algorithms are based on optimization targets. The existing target systems of production logistics need to be enlarged by new criteria so that in future products still can be manufactured feasible. Such criteria can be energy costs, flexibility or fuzziness in time.

Disturbances are part of every production environment. The reasons for deviations between order planning and realized production plan are manifold. Apart from simple machine failures these can comprise delays in order release due to late delivery of goods or restricted personnel capacity. Nevertheless, not every deviation should trigger a complete redesign of the production program. On the one hand, some deviations have just a marginal influence to the feasibility of the overall production. On the other hand, frequently changing machine usage plans would cause continuous flow of materials, equipment and resources between machines with extensive handling efforts. Consequently the development of a *robust control concept* is necessary that permits appropriate interventions but tolerates organic process divergences.

Basic master data of the specific operations are stored at ERP level. Over time, these information can differ from the real parameters at the shop floor. Apart from fluctuating material characteristics or changing operating procedures employee's influence can be reasonable for that. If target values are departing more and more from the actual situation, lead times or forecasted load peaks may be exceeded. Because of this, an integrated adjustment of master data at ERP level based on feedback loops from the manufacturing side should be striven to ensure appropriate schedule planning runs in the future.

4.1.2 Concept and Procedure

To solve these challenges, interdisciplinary research is needed. At Fraunhofer IPT flexible concepts for production planning and control are developed and implemented in a prototypical framework. These concepts integrate multidimensional targets like energy consumption, costs and time risks. For that, comprehensive transparency and accessibility of all necessary production data is crucial and realized by interlinking IT-systems at ERP, MDC, PDA, ECS and MES level.

The energy turnaround will cause significant changes at the German energy market. Apart from increasing energy prices innovative measures to guarantee stability of the electricity grid. On the way to smart grids network operators and power suppliers are thinking about flexible tariffs (e.g. time-of-use) and new load management structures to cope with erratic feed from renewable sources. In future it can be assumed that such programs will be more and more provided to manufacturing companies. Apart from threatening fines for load peaks this changing market structure will put the companies into a position of reducing their energy costs by getting bonus payments or lower prices for electricity. Though, such tariffs imply the company's ability to forecast, collect and influence their energy demand.

To enable especially small and medium sized companies to actively take part, the research project "eMES" (FKZ: 01IS14025A-D), funded by "Bundesministerium for Bildung und Forschung" (BMBF) was drafted. In cooperation with two industrial partners (software development and sensors) the concept will be implemented at the factory side of a manufacturing company. Main research target is the development of a prototype for an energy oriented production planning and control that is modular integrated in an existing MES. The MES creates an intra-day production plan and intervenes according to an energy control loop in case of deviations (see Fig. 5). Apart from machine failures such deviations can also be real

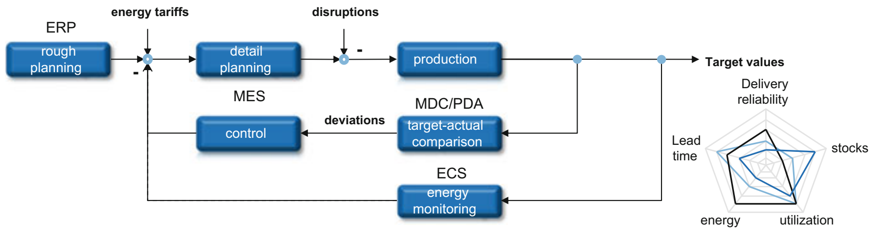


Fig. 5 Energy oriented control concept

load profiles that are reported back by the ECS to the control station and that differ from the modelled profiles significantly. Possible adjustments to the planning scenario are analyzed regarding their economic benefits and if necessary rejected.

For example, with these features companies will be able to assess the feasibility of shifting energy intensive process steps to off-peak periods. The deliverables of the project include, apart from the IT functionalities, also the definition and implementation of interfaces between sensors, machines and database connection to ERP- and ECS-systems. The key challenge for the practical implementation of the prototype particularly is in the integration of the existing machine control.

To cover the current status of production management in most companies a multistage and scalable planning tool is developed. This tool sorts the available operations to the machines via automated algorithm, easier sort rules or by manual scheduling (Fig. 6). This takes up the above-mentioned stepwise introduction of CPS and enables companies a gradual transformation of their experience-based planning processes. Summarizing that, in all planning alternatives the expected cost effects of the examined production plan are calculated and aggregated in form of an objective value with dimensions like machine costs, utilization, energy costs, delay costs or storage cost. Thus the comparability of the different scenarios is always ensured.

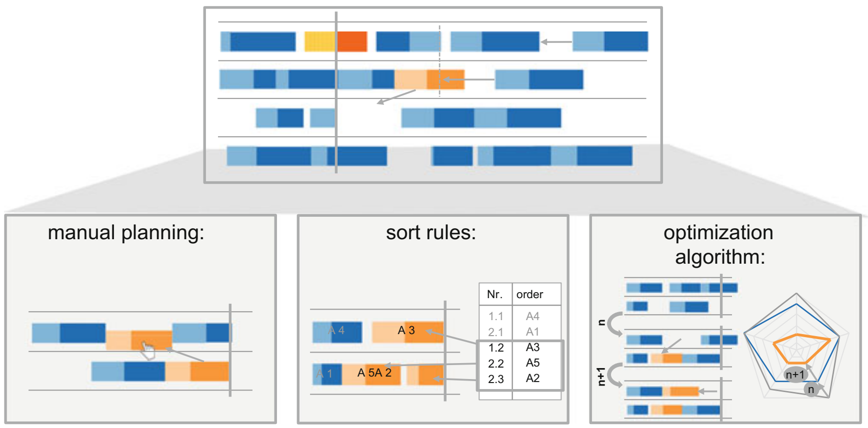


Fig. 6 Planning alternatives

After that, the chosen machine usage plan will be transferred into the production and executed. If deviations between target and actual values of the load profile or in case of error messages the control concept will actively decide if a redesign of the plan is useful at the current timeline. From an energetic point of view several aspects trigger that decision. Firstly, the relative difference between targeted and actual energy consumption and secondly the forecasted effect on the 15 min mean value as well as distance to the planned overall load peak is considered. In the validation case of the research project the final decision of the re-planning action is taken by human, but otherwise can be automatized easily. Thus such a solution would autonomously develop its optimal production plan and continuously proof this decision based on the reported information.

The resulting requirements regarding the production environments are mainly cross-linked and adaptive processes that adjust flexible to changing framework conditions and are able to work with short term modifications. Load peaks are determined by the maximum value of all the averages of 15 min periods. To react adequately to this energy cost driver, *short reaction times and a sufficient granularity of production data* (e.g. minute-by-minute precision of energy consumption) are mandatory. This leads to an analyzation of the production areas that are suitable for direct intervention or have the potential for load shifting. If not done yet these have to be equipped with electric meters and data loggers.

On the other side complete information from ERP side are inevitable for a reliable planning quality of the production program. From an energetic point of view, this comprises apart from fundamentally information like delivery date, processing times and resources also the product specific energy consumption of the process step. In times of increasing variants and product diversity the point of setting up and keep this master data up to date is more and more challenging. At least when there are hundreds of products, each comprising of a number of operations, this can't be done manually anymore. To solve this, the procedure of initializing and updating the information needs to be done internally and automated in the MES. For example the operation related energy profiles can be drafted by linking the time-continuous measured energy data to the status information from the shop floor.

Coming from the theoretical concept, two main restrictions has to be soften in practical implementation. On the one hand the operation-related energy consumption that characterizes the every minute power input of the machine during the manufacturing process, is often significantly more variable. Based on the evaluation of real energy data from the validation partner the average power can vary up to 50 %, depending on type, characteristics of the raw material and user influence. On the other hand -due to the same reasons- the processing time is also flexible in certain ranges. Dealing with such scattering processes robust control concepts are needed accordingly. In this case, robust means that deviations, even if they are considerable, are tolerated as long as there are no cost effects or changes in the target figures.

4.1.3 Applications and Potentials

Because of its scalable functional range the developed concept for production planning and control can be implemented in basically every company. Nevertheless if the company is already collecting energy data and having a maintained ERP-database, the implementation effort would be reduced drastically. Besides, energy should have a relevant share in the overall cost structure of the company and the production processes have sufficient load shift potential and allow short term interventions. Depending on the company specific targets the application can range from enhancing transparency of the several processing steps (monitoring) to autonomous operation of an entire factory, where the control station manages machines independently.

Whilst in case of monitoring primarily process stability and due to short-term adjustments constant product quality is assured, the usage of optimization algorithm leads to a cost efficient production flow. Early adaptions as an effect of deviations and disturbances reduce waiting times and improve machine utilization. This directly cuts production costs. Additionally—depending on tariff structure—energy costs can be reduced due to lower performance prices derived from the effective load peak, intelligent shifting of operations in times with cheaper electricity prices or by getting bonus payments owing to providing of load shedding capacities. According to first simulations the usage of the developed concept has the ability to reduce load peaks up to 30 % regularly.

Secondary effects are expected by analyzing energy data of the several process steps. With comparison of load profiles of the same product over the time, reliable recommendations regarding tool life and failures can be derived ("predictive maintenance"). The research project eMES lays first foundations for implementation of CPPS. Because of interlinking of IT-structures, comprehensive production data acquisition, implementation of control loops and consideration of multi-criteria target values, the production environment is trained for an ongoing autonomization.

4.2 Smart Glasses in Industrial Assembly

The overall goals for the usage and analysis of production-related data often focus on quality, productivity and resource efficiency as well as the reduction of waste. In order to achieve these goals, employees can be supported through the targeted usage of data. In this case not only the generation of information is necessary, but also its quick availability and its utilization on local workstations need to be guaranteed to enable workable CPPS-structures and an ideal real-time support.

In technical terms, this availability can be realized through so called Smart Devices. These are generally mobile devices like tablet computers, smart phones and data glasses which, as a link between the physical and virtual world, can provide production employees with filtered and specific information during their work. Thus it enables employees to take decisions and optimize their local working process on their own. The mobility of those devices and therefore location-independent access obtain a reduction of processing time in contrast to stationary devices, like e.g. terminals. Next to an appropriate IT-infrastructure, which e.g. enables data transfer via WLAN, it has to be possible to integrate the device into existing IT systems to reduce data interfaces. The device must be able to interact with such systems and the user. This ability combined with the (at least limited) decentral and autonomous data processing ability describes the device to be "smart". The delivery of relevant information, which can support the user in decision finding or even generation of decision situations, is associated with the approach of distributed production control [5]. The employee takes over the decision of the central production control system and solves problems without the intervention of it.

In general, Smart Devices are unlimited in matters of their informational direction. They can receive and process data to inform the employee but also gather and edit data to store it in linked IT systems like MES, ERP or CAQ. To provide the devices with the relevant functions, these can be equipped with buttons, touch-screen, cameras, temperature, and acceleration sensors. If Smart Devises are solemnly used for data acquisition, valuable potential will be neglected because the experience and decisions of an employee will not be considered. In future implementations it is expected that Smart Devices support production employees while simultaneously dealing with documentation, information, communication and measuring tasks.

Due to the fact that users generally carry or even wear Smart Devices and that those devices can perceive their surrounding through their sensors, these are also suitable to implement novel occupational safety systems. Thus, a Smart Device can support the localization of employees or e.g. inform an open access robot system through interfaces about risk of collision. Next to an application as measuring device to realize an efficient self-checking by the employee and fast preliminary decisions, new business models can be derived considering various different interfaces, which are available even today.

4.2.1 Overview of Smart Devices Established on the Market

As mentioned above, the collection and provision of process and product information at local workstations is crucial to facilitate CPPS structures. The usage of smart devices will help companies to create such conditions with minimal efforts. In addition to smart phones and tablet computers, more and more technologies or devices in daily life are equipped with "smart" functions. To add "Smartness" to well-known and often used goods promises a higher user acceptance. Especially Smart Wearables, like Data Glasses, seem to be a suitable solution as users are already used to wear similar clothes or accessories. Considering current trends it becomes obvious that generally Smart Devices are developed which are compatible with habits of the users.

Smart Wearables are still mainly developed for private use. The largest economical potential of Smart Devices however, according to a new study, is in industrial and production applications [30]. Thus these new developments will increase in producing companies.

In comparison to other Smart Wearables, Smart Watches, Wristbands or Glasses appear specifically favorable because these solutions can transfer a large amount of information through visual interfaces. Here, the usage of displays is crucial. Simply by an appropriate arm positioning, the user has access to the content and information on the display of Smart Watches and Wristbands. Smart Glasses are already in the right position simply by putting them on. Whereas Smart Watches and Wristbands generally only use one display, Smart Glasses can use one display per eye. Additionally, displays can be distinguished in transparent and closed displays. In addition to the application of different display technologies, development effort has been put into the direct projection of relevant pictures on the retina to improve the comfort of reading. The producer Fujitsu aims at a focused image which is independent of focusing of the eye [18]. Market maturity of this technology is not achieved yet. The most popular Smart Glass manufactures on the current market are Samsung, Apple and Sony. Other manufactures are e.g. Vuzix, Zeiss and Google.

In addition to the introduced display screen, Smart Devices also can transfer information to the user per light signal, vibration or sound output. The common sensors are cameras, microphones, acceleration, inclination sensors but also vital sensors, like heart rate sensors. The latter offers new possibilities in the field of Occupational Security. Especially in security-critical work environment the vital conditions of a person can be decisive for the own life or the life of others. Cameras and microphones can be used for communication or documentation purposes. In connection with acceleration and inclination sensors, cameras can be utilized to implement Augmented Reality functions to the Smart Glasses. In consideration of the captured Glasses' movements virtual objects can be placed in the real world. Through this, goods can be marked specifically, e.g. at commissioning processes, to minimize error rates in intralogistics.

4.2.2 Current Applications and Suitability of Smart Devices in Production

Currently only few Smart Wearables are suitable for an industrial application. The main reasons for this are the battery life and the robustness, among others. Thus, the current devices are rarely able to cope with extreme conditions like dust, moisture, high and low temperatures. For manufacturing companies, however, these characteristics are essential, because otherwise excessive downtime and repair costs can be expected. Also the currently built-in processing power and storage capacity is not sufficient for many applications [19]. For industrial usage, the glasses of Vuzix have been identified by Fraunhofer IPT as one of the products with the highest application readiness and have thus been used in pilot tests after all glasses available on the marked have been compared in a benchmark.

In addition to hardware constraints, the software portfolio for industrial applications is particularly low, as developers have initially focused on private use and thus smartphones and tablets. Complex tasks must therefore still be dealt with by using customized software solutions. The results of the study discussed above (see also [30]), however, indicate a future growth of industrial software offers. Companies which want to optimize production and implement CPPS in mid-term or long-term with new technologies should now start to deal with the issue of Smart Devices. For that, it is necessary to define applications and requirements. Former industrial and research projects show that alliances with established software vendors achieved the highest impact.

4.2.3 Potentials for Increased Productivity and New Applications

Smart and mobile devices offer great potentials in terms of reduced process times, quality and resource efficiency. In addition to that, they support the operational safety and even offer ways to new business models. As mobile and smart measurement and documentation systems, they depict a significant contribution to the digitization of the production within the scope of Industry 4.0. As part of a research initiative, support functions in processes, such as testing and assembly processes, have been implemented within the Fraunhofer IPT (see Fig. 7). Here, employees received instructions as combinations of text, image and video descriptions to assist him in the execution of complex tasks. By voice commands or optionally by buttons, shop floor employees and test persons were able to navigate through the software. In the future, also gesture recognition should be realized through the glasses' camera. The use of voice recognition offers the advantage that users can interact with Smart Glasses without using their hands.

To plan the assistance of the Smart Glasses, the entire process needs to be modeled using a self-developed, browser-based application. Using an intuitive interface which can be accessed via Smart Phone, tablet or desktop computer, users have a barrier-free possibility of process modeling. These processes will then be sent

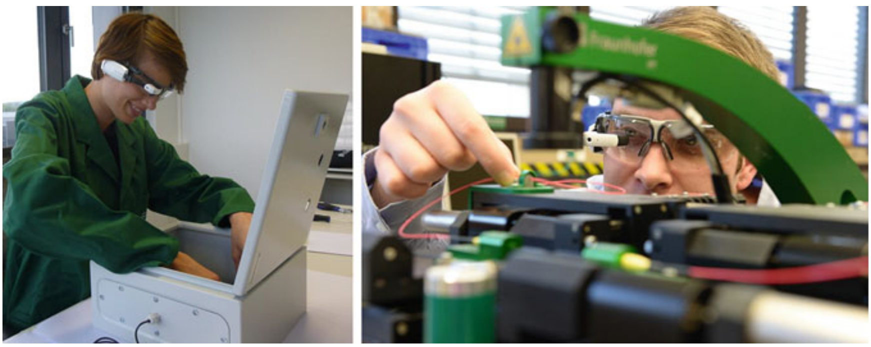


Fig. 7 Smart glasses as CPAS in assembly tasks

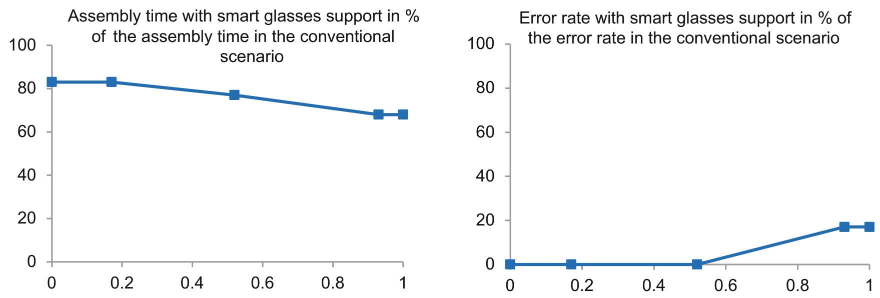


Fig. 8 Reduction of lead time and failure rate by the application of smart glasses as CPAS

to the Smart Glasses via WLAN. The user registers itself with a QR code that is detected by the glasses' camera. Further applications of the QR code recognition, such as querying product data in picking processes, are subject to current research. Through a comprehensive study by Fraunhofer IPT, the reduction of errors in an assembly scenario as well as the average reduction of assembly time could be verified (Fig. 8). An important reason is that test persons can see the notes in parallel during the assembly process and thus, compared to conventional operating procedures on paper, mental set-up processes can be reduced. With increase of complexity of the investigated assembly process the lead-time is decreasing. The process complexity was especially operationalized with a variants-induced scale from 0 to 1.

In this context, the usage of Smart Glasses for product assembly can be integrated in Cyber-Physical production systems as a assisting functionality that supports the employees to reduce product failures and optimize their productivity.

5 Prerequisites and Requirements for Cyber-Physical Production Systems in the Resilient Factory

The scientific and technological as well as social conditions for the successful launch of Cyber-Physical production systems are extremely diverse. Among others, the shift towards decentralized IT structures, the creation of new approaches to data and communication security, the modularity of technical systems, the changes in the network structure of manufacturing companies and last but not least the successful integration of the people in the socio-technical system "factory" are essential conditions (see [9, 25]).

The technical and scientific requirements for Resilient Factories in the context of Industry 4.0 are discussed in the following chapter. These conditions can be hardly established by only a few companies as technology leaders and, in the context of horizontal networking over many process chains, not in all companies simultaneously. These are the main reasons why the way to Industry 4.0 is understood as an evolutionary process taking one step at a time.

5.1 Technical Requirements

5.1.1 Decentralized IT Structures

An essential basic concept of Cyber-Physical systems is the shift away from classic hierarchical structures towards versatile, adaptive networks of (production) entities (see e.g. [25]) (see Fig. 9). This change causes diverse conditions and challenges for the future design of production systems: Systems must be able to communicate via open standards and protocols. Negotiations and cooperation to achieve goals are essential behaviors of CPS. For this purpose, services are offered and perceived within the network. Furthermore, the self-description of systems is an essential prerequisite in order to identify possible negotiation and cooperation partners and which collaborations and services are possible [9].

Similar approaches of a so-called agent-based production control as implementations of decentralized controlled networks have been explored since the 1980s (see [34]). There are individual entities of a network competing on a market place to get jobs, they cooperate and compete against each other. The decentralized approach proved to be extremely robust to external disturbances and was also able to provide near-optimal results for complex problems of production management and control in an acceptable computation time (see [24, 31]). In this case, however, these projects were pure research projects and that is why many practical authors see this approach with some skepticism (see e.g. [21]). So it is noticeable that even in new sources, although there are many theories and concepts, only little real use cases in production environment are mentioned (see [20, 24]).

If production systems are not longer guided by a central controller, further challenges in the field of planning arise: Transparency of the overall system condition, late contracts and related delivery times are significantly more difficult to obtain in a decentralized network. Also reschedules of individual orders and process, changes cannot longer be managed by a central control instance. They e.g. have to be managed by the network through prioritization and virtual prices [39].

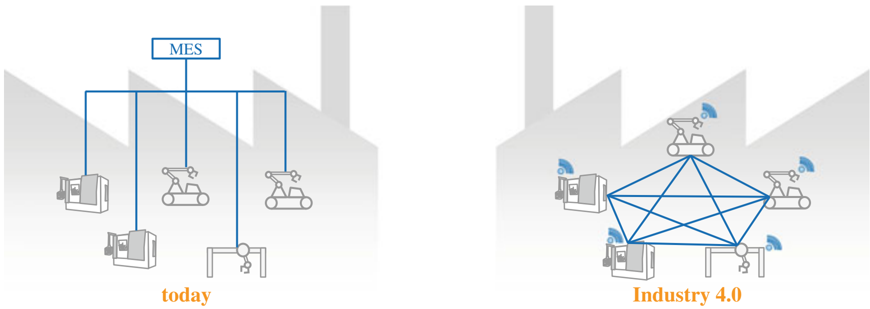


Fig. 9 Change from hierarchical to decentralized systems in an industry 4.0 environment

5.1.2 Data Security

The protection of data and know-how as well as the protection against external attacks is much more difficult to achieve in decentralized, adaptive networks. Security is therefore an essential requirement for the successful implementation of cyber-physical systems in production [4]. Communication thus needs to take place primarily between systems "trusting each other" in the sense of certificate exchange etc., as trust is associated with greater safety. Identity, integrity and encryption are thus properties which will be essential for the security of networks.

5.1.3 Modularity of Technical Systems

Closely related to the decentralization of IT structures is the demand for greater modularity in the Cyber-Physical Systems. Layer, or even cross-company networking can occur only when standardization allows a modular structure of the entire system. In addition to extensive communication standards and interfaces, this especially requires reference architectures, i.e. an overall definition of mechanisms for collaboration and the exchange of information (see [25]).

In the past, many scientific projects addressed the modular design of production systems. Using the expression "Plug & Produce", many publications and current projects can be found (see e.g. [33] or [35]). Whether such an openness and standardization of interfaces can be achieved in comparison to existing, highly individual and protected interfaces, e.g. in the programming of industrial robots,, remains subject to further research and market development.

5.1.4 Fusion of Shopfloor- and Office-IT

In most manufacturing companies today there is a sharp separation between IT systems in the office environment (white collar world) and the shop floor (blue collar world). Depending on the company structure, this is even reflected in different supervising technical departments: The automation of production is carried out by SPSS and appropriately trained technicians, while the care of computers and systems on the planning side is done by appropriate system administrators (see Fig. 10).

However, the fusion of these structures is an essential precondition for the successful introduction of Cyber-Physical Systems to manufacturing companies. Real vertical integration can only be permanently achieved if data from the (deepest) machine-level to the super ordinate levels of management, is transferred, aggregated and processed. But the reverse way is also to be noted here: Changing conditions such as prices for resources, changing delivery times or customer requirements are placed on various (management) levels in the system, but can have an impact to the machine level.

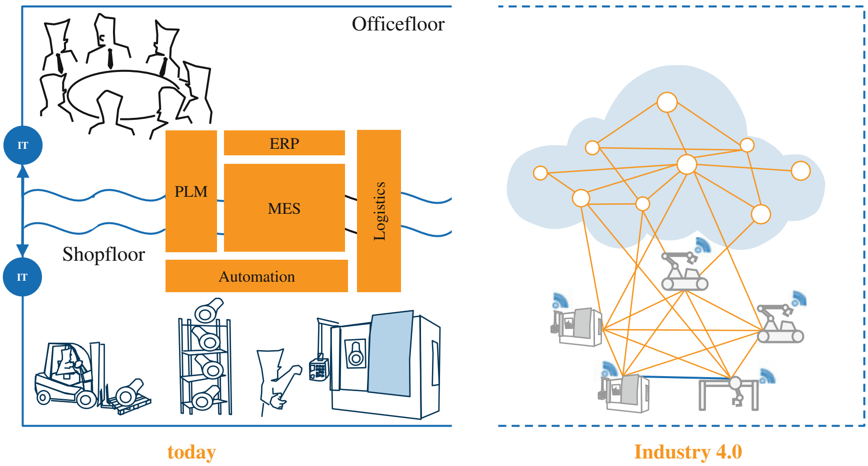


Fig. 10 Fusion of IT-systems in the blue and the white collar world

In addition, the future network in manufacturing companies must be perceived as a wearing component of the overall system and treated accordingly. Robustness is possible only if a regular review of the technical condition of the existing data connections, their utilization and performance is done.

5.1.5 Integration of People: Socio-Technical Systems

Another essential requirement for the successful introduction and implementation of Cyber-Physical systems in production is the integration of employees into a socio-technical system. This includes requirements for both the technology as well as to the qualifications and skills of the employee himself.

On the technical side, the operational safety needs to be guaranteed under all conditions (see e.g. [4]). Furthermore, the ability to self-description and clear presentation of the system status and the conformant behavior depict necessary conditions [9]. Conformant behavior matters most in areas of direct co-operation between human and machine, such as industrial assembly. This necessities, for example, that movements of the system as well as its response to outer conditions are in accordance to human expectations, implying that it is designed anthropomorphic [7].

In the future, comprehensive education must take precedence over technical specialist training, because of the close integration of production and IT [25]. Also the employee in production will meet new challenges with adaptive, IT-heavy production systems, for example, placed in the field of diagnosis of conflicts and disturbances, the planning or the cooperation of technical components.

5.2 Scientific Requirements

As shown in the last chapter, several technical requirements still need to be met to be able to successfully introduce Cyber-Physical Systems into a production environment. In addition to that, a multitude of open questions and even risks need to be answered from a scientific stand point. These are mainly questions resulting from their major characteristics: their ability to learn and adapt, their decentralized structure as well as the self-descriptiveness in a network of interacting systems. These can only be answered in a close cooperation between scientific research and industrial application to ensure the quick introduction of new concepts, methods and tools into the technical state of the art that was already discussed in the preceding chapter.

5.2.1 Reliability of Adaptive, Learning Systems

Systems that react on environmental changes and thus alter their behavior need to be monitored and controlled closely to prevent undesirable developments. Cognitive structures as e.g. SOAR (compare [27]) are using feedback and rewards to evaluate alternative problem solving strategies and thus learn new patterns. This approach can contribute to a much faster and more robust control of complex systems (e.g. [37]). But the evaluation of alternatives typically requires the comparison of differing, sometimes even contradicting metrics and goal values [38]. This may provoke the change within an adaptive, learning system that appeared feasible from a mathematical stand point, but instead caused harm economically due to a false interpretation or maybe even negligence of certain boundary conditions and causational relationships. These need to be recognized and modelled in advance.

5.2.2 Integration of Risk Management for Non-deterministic Systems

Due to ability of Cyber-Physical Systems to learn and adapt to changes in the environment as well as their decentralized cooperation, an a prio assessment of their expected behavior is almost impossible. This may lead to inconsistencies and fuzziness on an organizational level (e.g. due dates, job routing) as well as on a technical process level (e.g. energy supply, updates, maintenance).

Risk management as a means of preventive quality management requires a clear depiction of dependencies, causal relationships and probabilities in order to assess and prioritize risks as well as their countermeasures. This appears almost impossible when looking at the non-deterministic behavior patterns of Cyber-Physical Systems discussed above. Thus, new tools and measures need to be developed to enable risk management as a means of preventive quality management for these systems and their industrial application.

5.2.3 Model-Based Cooperation: Contradictions, Incompleteness, Failures

Models are usually applied to enable self-descriptiveness in networks of cooperating Cyber-Physical Systems. They are regularly applied to allow for planning and controlling of processes on different levels of value creation (compare [46]). Depending on the type and design of the specific models, they will exhibit deviations from the real world. They could be incomplete or, in an utmost extreme, even be in contradiction with the sensor-values giving the real-world representation of the current environment status. Major scientific questions will thus be the handling of failures, contradictions and incompleteness in predictions and models as well as their fast adaption to current needs. Examples of quick adaptions of process models by means of meta-models [2] and re-training of artificial neural networks [45] have been developed in the course of the Cluster of Excellence.

5.2.4 Emergent Patters

Emergent patterns usually occur in the genesis of objects or structures of smaller, fundamental entities [16]. They are consequences of the synergistic interactions of these entities, whether by pure coincidence or planned cooperation, which could not be expected or assessed when simply looking at the singular parts. The first descriptions of emergent patterns originated from swarm theory [16] as well as economic theories of capital market structures [15]. Emergent patterns can also be found in decentralized, agent-based production networks [3]. Their probability of appearance as well as consequences are thus of high scientific interest for Cyber-Physical Systems, as they are organized in a comparable way. Major areas of work are influence and interference of entities in networks as well as hampering, competition, even faster learning from others and information spread in networks.

5.2.5 Conflicting Data and Information

Cyber Physical Production Systems receive information about their environment from their own sensor as well as through the internet. Sensors thus depict one of the main components for the development of CPPS [13]. The availability of information will lead to mathematical overdeterminacy. If these deviate from one another due to sensor defects or disturbances, a large potential for conflicts arises. The right fusion of sensor data and information from the network will thus depict one of the major scientific questions for a successful introduction of Cyber-Physical Systems in a production environment.

6 Conclusion

The introduction of Cyber-Physical Systems into a production environment will foster the implementation of resilient factories, but at the same time create fundamental changes in contrast to today modus operandi. On the one side, this implies major changes for the structure and organization of producing companies and on the other side a strong penetration of machines and systems with sensors as well as elements for information processing and networking. These of course foster new technical risks and at the same time pose new requirements to the qualification and abilities at the shop floor. In contrast to these challenges, Cyber-Physical Systems create opportunities with regard to flexibility and adaptability as well as energy- and resource efficiency by mobilizing unused potential and thus create a higher productivity in spite of a rising variability.

Key element for the successful implementation of Cyber-Physical Systems in a production environment is the ensuring of the factories robustness. Adaptability, robustness, energy- and resource efficiency thus define the core characteristics of future production systems and thus together form resiliency in the context of Industry 4.0.

Companies have to face several challenges occurring from the introduction of Cyber-Physical Systems. IT-systems will become decentralized in the near future, dissolving the paradigm of centralized production control in hierarchical systems. Instead, production systems will be organized in decentralized networks. Security as defined by safe data exchange and communication requires identity, integrity and encryption in the entire network. To achieve collaboration across different machines and processes, modularity becomes a key element, as it is already addressed today by means of "plug and produce". To successfully implement the necessary changes in IT networks and infrastructure, the existing separation between IT-systems in the office and the shop floor level needs to be overcome. As humans will still depict the center of the future production systems, their integration into a more digitalized production environment needs to be tackled: Especially in the education of apprentices as well as engineers, aspects of information technology need to be integrated as basic knowledge.

From a scientific view, certain aspects need further consideration in the context of the introduction of Cyber-Physical systems into a production environment. Major questions will be the integration and control of as well as the risk management for non-deterministic systems: Understanding, assessing and even predicting the behavior of learning, adapting Cyber-Physical Systems showing emergent patterns.

In contrast to their overall impact, Cyber-Physical Systems are not expected to form the next industrial revolution in the sense of a dramatic change in a very short time. Instead, a step-wise evolution and permanent convergence is expected. As the discussed application examples have shown, Cyber-Physical Production Systems already exist today. The major questions, that this essay tried to assess, is thus not if, but how the fourth industrial revolution will take place and how companies as well as scientific research can foster its implementation to realize its utmost potential.

References

1. Adelt P (2009) Selbstoptimierende Systeme des Maschinenbaus: Definitionen, Anwendungen, Konzepte, 1. Aufl. HNI-Verlagsschriftenreihe, Bd. 234. Heinz-Nixdorf-Inst, Paderborn
2. Auerbach T, Beckers M, Buchholz G, Eppelt U, Gloy Y, Fritz P, Al Khawli T, Kratz S, Lose J, Molitor T et al (2011) Meta-modeling for manufacturing processes. In: Jeschke S, Liu H, Schilberg D (eds) Intelligent robotics and applications. Springer, Heidelberg, pp 199–209
3. Barbosa J, Leitao P (2011) Simulation of multi-agent manufacturing systems using agent-based modelling platforms. In: 2011 9th IEEE international conference on industrial informatics (INDIN), pp 477–482
4. Bettenhausen D, Kowalewski S (2013) Cyber-Physical Systems: Chancen und Nutzen aus Sicht der Automation, Düsseldorf
5. Botthof A, Hartmann E (2015) Zukunft der Arbeit in Industrie 4.0. Springer Vieweg, Berlin
6. Brecher C (ed) (2011) Integrative Produktionstechnik für Hochlohnländer. VDI-Buch, Springer, Berlin
7. Brecher C, Müller S, Kuz S, Lohse W (2013) Towards anthropomorphic movements for industrial robots. In: Duffy V (ed) Digital human modeling and applications in health, safety, ergonomics, and risk management. Human body modeling and ergonomics, vol 8026. Springer, Berlin, pp 10–19
8. Broy M (2010) Cyber-Physical Systems: Innovation Durch Software-Intensive Eingebettete Systeme. Acatech DISKUTIERT, vol 0. Springer, Berlin
9. Broy M, Geisberger E (2012) AgendaCPS: Integrierte Forschungsagenda; Cyber-Physical Systems. acatech-Studie [1]. Springer, Berlin
10. Brunner FJ (2008) Japanische Erfolgskonzepte: Kaizen, KVP, Lean Production Management, Total Productive Maintenance, Shopfloor Management, Toyota Production Management. Praxisreihe Qualitätswissen. Hanser, München
11. Bullinger H, Hompel M (2007) Internet der Dinge. www.internet-der-dinge.de. VDI, Springer, Berlin [u.a.]
12. Bundesministerium für Bildung und Forschung (2012) Zukunftsprojekte der Hightech-Strategie (HTS-Aktionsplan). Bericht der Bundesregierung, Bonn
13. Bundesministerium für Wirtschaft und Technologie (2013) Multimodale Sensorik-Konzepte der Umwelt erkennung/-modellierung. Leitfaden im Rahmen des Technologieprogramms Autonomik, Berlin
14. Carvalho H, Cruz Machado V (2006) Fuzzy set theory to establish resilient production systems. In: IIE Annual Conference
15. Chen S-H, Yeh C-H (2002) On the emergent properties of artificial stock markets: the efficient market hypothesis and the rational expectations hypothesis. J Econ Behav Organ 49(2):217–239. doi:10.1016/S0167-2681(02)00068-9
16. Cucker F, Smale S (2007) Emergent behavior in flocks. IEEE Trans Autom Control 52 (5):852–862
17. CyProS Produktivitäts- und Flexibilitätssteigerung durch die Vernetzung intelligenter Systeme in der Fabrik (2015). www.projekt-cypros.de
18. Das Auge als Bildschirm: Datenbrille lasert direkt auf die Netzhaut (2015) Deutsche Wirtschaftsnachrichten
19. Dlugosch G (2015) Datenbrille bringt Durchblick im Lager. VDI Nachrichten
20. Dumke R, Mencke S, Wille C (2010) Quality assurance of agent-based and self-managed systems. CRC Press, Boca Raton
21. Eversheim W, Schuh G (2014) Produktion und management. Springer, [S.l.]
22. Gausemeier J (2009) Selbstoptimierende Systeme des Maschinenbaus, Definitionen und Konzepte. Westfalia Druck, Paderborn
23. Hollnagel E, Woods DD, Leveson N (2006) Resilience engineering: concepts and precepts. Ashgate, Aldershot

24. Jennings NR, Wooldridge MJ (1998) Agent technology: foundations, applications, and markets. Springer, Berlin
25. Kagermann H (2013) Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0: Abschlussbericht des Arbeitskreises Industrie 4.0, Frankfurt (Main)
26. Kamiske GF, Brauer J (2011) Qualitätsmanagement von A–Z: Wichtige Begriffe des Qualitätsmanagements und ihre Bedeutung, 7, aktualisierte Auflage. Hanser, Carl, München
27. Laird J (2012) The Soar cognitive architecture. MIT Press, Cambridge
28. Lee EA (2013) Cyber-Physical systems: a rehash or a new intellectual challenge. Distinguished speaker series, Texas
29. Martin-Breen P, Anderies JM (2011) Resilience: a literature review. Rockefeller Foundation, New York
30. McKinsey Global Institute (2015) The Internet of Things: mapping the value beyond the hype
31. Monostori L, Váncza J, Kumara SRT (2006) Agent-based systems for manufacturing. CIRP Ann Manuf Technol 55(2):697–720
32. Nyhuis P et al (2008) Wandlungsfähige Produktionssysteme. Heute die Industrie von morgen gestalten. PZH Produktionstechnisches Zentrum, Hannover
33. Onori M, Lohse N, Barata J, Hanisch C (2012) The IDEAS project: plug & produce at shop-floor level. Assembly Autom 32(2):124–134
34. Parunak HVD (1988) Manufacturing experience with the contract net. In: Huhns MN (ed) Distributed artificial intelligence. Morgan Kaufmann, California
35. Reinhard G, Krug S (2012) Automatic configuration (plug & produce) of robot systems—data-interpretation and exchange. In: Enabling manufacturing competitiveness and economic sustainability: proceedings of the 4th international conference on changeable, agile, reconfigurable and virtual production, pp 147–152
36. Schmitt R (2011) Selbstoptimierende Produktionssysteme. In: Brecher C (ed) Integrative Produktionstechnik für Hochlohnländer. Springer, Berlin, pp 747–1057
37. Schmitt, R et al (2010) Cognition-based self-optimisation of an automotive rear-axle-drive production process. J Mach Eng 10(3):68–77
38. Schmitt R, Permin E (2013) Characterization of future key performance indicators and constraints for self-optimizing assembly systems. In: Schmitt R (ed) Proceedings of the 11th international symposium on measurement technologies and intelligent instruments. Shaker, Aachen
39. Schmitt R, Permin E, Losse S (2014) Achieving resource- and energy-efficient system optima for production chains using cognitive self-optimization. In: Selinger G (ed) Proceedings of the 11th global conference on sustainable manufacturing, pp 570–575
40. Sedlacek K (2010) Emergenz: Strukturen der Selbstorganisation in Natur und Technik: Arbeits- und Handbuch, 1st edn. Books on Demand, Norderstedt
41. Sheffi Y (2005) The resilient enterprise: overcoming vulnerability for competitive advantage. MIT Press, Cambridge
42. Shingō S (1996) Quick changeover for operators: the SMED system., Shopfloor seriesThe Press, Portland
43. Spath D, Ganschar O (2013) Produktionsarbeit der Zukunft: Industrie 4.0, Stuttgart
44. Vogel-Heuser B, Pohl J, Hausmanns C, Michel B (2015) Alles neu macht Industrie 4.0. In: Produktion 11/2015
45. Wagels C, Schmitt R (2012) Benchmarking of methods and instruments for self-optimization in future production systems. Procedia CIRP 3:161–166
46. Wea S (2011) Virtuelle Produktionssysteme. In: Brecher C (ed) Integrative Produktionstechnik für Hochlohnländer. Springer, Berlin, pp 257–464
47. Westkämper E (2006) Einführung in die Organisation der Produktion, 1st edn. Springer-Lehrbuch, Springer, Berlin

Part IV
Communication and Networking

Communication and Networking for the Industrial Internet of Things

Jan Rüth, Florian Schmidt, Martin Serror, Klaus Wehrle and Torsten Zimmermann

1 Introduction

Recently, a vast area of research emerged, known as the Internet of Things (IoT), aiming at connecting all sorts of everyday devices to the Internet to enable new ways of interaction and automation [5]. Besides academia, the industry has a great interest in exploiting the potential of the IoT, e.g., in developing new products, which make their services remotely accessible to humans as well as to other devices. Primarily, the IoT centers around human beings striving constantly to improve their daily life through remote accessibility, the interconnection of services and tangible interaction [77]. A loosely coupled sub-area of the IoT is the *Industrial* Internet of Things (IIoT), which, in contrast, puts the focus on the manufacturing process in industrial automation, i.e., adding value to it by globally connecting machine parts, production entities, and factories [13].

Similar to the IoT, the IIoT builds upon the premise that a globally accessible communication infrastructure is available to a plethora of devices involved in industrial processes. However, in the past, communication within factory halls was

J. Rüth · F. Schmidt · M. Serror (✉) · K. Wehrle · T. Zimmermann (✉)
Communication and Distributed Systems, RWTH Aachen University,
Ahornstr. 55, 52074 Aachen, Germany
e-mail: serror@comsys.rwth-aachen.de

T. Zimmermann
e-mail: tzimmermann@comsys.rwth-aachen.de

J. Rüth
e-mail: rueth@comsys.rwth-aachen.de

F. Schmidt
e-mail: schmidt@comsys.rwth-aachen.de

K. Wehrle
e-mail: wehrle@comsys.rwth-aachen.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_12

rather simple, locally connecting sensors, actuators, and controllers for specific automation tasks [46]. In the end, an interconnection of devices on the factory hall level and even between factories scattered around the world in combination with open standards will allow for a more precise synchronization within and between production facilities. Moreover, it will facilitate just-in-time manufacturing, as production processes may be altered and adapted locally and globally.

The introduction of wireless technologies for locally connected production entities introduces high flexibility to industrial processes, significantly reducing the costs for deployment and maintenance, as cables are negatively affected by wear and tear, and exposition to harsh industrial environments [21]. However, current standards for wireless communications, such as the IEEE 802.11 standard family [33] or Bluetooth [32], are not suited for industrial automation, as they are not able to provide the needed communication guarantees both in latency and reliability. It is thus necessary to develop new standards, targeting the aforementioned guarantees while, at the same time, providing interoperability to other services envisioned in the IIoT that do not rely on such stringent communication guarantees.

This chapter aims at providing an overview of existing communication technologies and how they may be adapted for the IIoT. Therefore, it addresses the unique challenges for implementing the IIoT, referring to its heterogeneous components. It is structured as follows. In Sect. 2, we review existing communication systems for industrial automation and the challenges that arise when implementing the IIoT. Then, in Sect. 3, we provide an overview of local communication within industrial automation and especially focus on the use of wireless technologies. Section 4 explains how this locally organized communication infrastructure can be extended to a global scope, enabling communicating within factory halls and beyond. On top, application layer communication, which is a vital part of the IIoT, is introduced in Sect. 5. Finally, a short conclusion and outlook is provided is Sect. 6.

2 Communication in Industrial Automation

This section provides an overview of the role of communication systems in industrial automation. After defining the major terms relevant for this chapter (cf. Sect. 2.1), follows a short introduction to current communication systems used in the domain of industrial automation (cf. Sect. 2.2). Finally, the section concludes with an identification of the goals and challenges that are required in the Industrial Internet of Things (IIoT) (cf. Sect. 2.3).

2.1 Definitions

In general, computer-integrated manufacturing systems comprise three distinct components: sensors, controllers, and actuators [23]. An overview of these components, as well as their mutual dependencies, is depicted in Fig. 1. Their responsibilities can be described as follows.

Sensors continuously or periodically measure a target parameter in their respective environment. Output values are either provided regularly with a certain interval or event-based, i.e., when a predefined threshold is exceeded. The unfiltered output values are typically conveyed to a controller.

Controllers regularly receive measurements from sensors as input. The measurements are stored, processed, and continually fed into a control function mapping the measurements to an output value, e.g., using a differential equation. This output value is then sent as a command to one or more actuators.

Actuators act upon commands received from a controller. Possible actions are, e.g., moving a machine part to a certain position, changing the rotation speed of an engine shaft, or performing an emergency stop. Actuators thus directly influence the environment in which they operate.

The ISA-SP100 working group defines six different application classes for industrial automation, involving sensors, actuators, and controllers, which can be categorized into three groups [36], see Table 1. Each class has its own requirements in terms of Quality-of-Service (QoS) towards the underlying communication system. In general, it can be stated that the lower the class, the more stringent are the guarantees required by the application regarding latency and reliability.

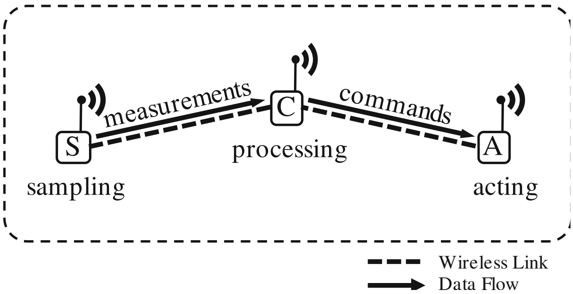


Fig. 1 Typical data flow in wireless industrial networks: Sensors (*S*) periodically send measurements to controllers (*C*), which, after processing the data, send commands to actuators (*A*)

Table 1 Overview of the ISA-SP100 application classes for industrial automation [36]

Category	Class	Application	Description
Safety	0	Emergency action	Always critical
Control	1	Closed-loop regulatory control	Often critical
	2	Closed-loop supervisory control	Usually non-critical
	3	Open-loop control	Human in the loop
Monitoring	4	Alerting	Short-term operational consequence
	5	Logging and downloading/uploading	No immediate operational consequence

In current automation systems, the controller is often implemented separately from sensors and actuators, centrally collecting input from various sensors and controlling different actuators, e.g., a Programmable Logic Controller (PLC). This implies that the considered components must be connected via a communication system to be able to exchange data. As stated before, some application classes require stringent communication guarantees to function properly. This includes, e.g., safety applications, where (small) data packets must be transmitted with a high reliability and a guaranteed low latency. Critical closed-loop control similarly imposes stringent guarantees on the reliability and the latency. These applications thus require adapted communication systems that are able to provide the needed guarantees. Typical target values lie in the area of a few milliseconds for latency and a packet loss rate below 10^{-9} for reliability [21]. Achieving these target values is particularly challenging with wireless transmission systems, as prevalent wireless technologies, such as IEEE 802.11 standards [33], were developed for home and office environments, optimizing throughput instead of latency and reliability. Moreover, harsh industrial environments may further deteriorate the wireless link quality due to path loss, shadowing, and multi-path propagation [74]. In contrast, other application classes, e.g., monitoring, have more relaxed requirements on the communication system and are consequently easier to support with wireless systems.

However, the future IIoT must able to support all six application classes, including wireless technologies, to provide comprehensive benefits for industrial automation. The most challenging ones are thereby the critical ones, i.e., Classes 0 and 1, which, nevertheless, are limited to local automation setups as further discussed in Sect. 2.3. In the next section, we review current communication technologies in industrial automation regarding their suitability for the IIoT.

2.2 *Current Trends in Industrial Communications*

In the last decades, communication systems for industrial automation were mostly designed to locally interconnect sensors, actuators, and controllers using Fieldbus technology or Ethernet [46]. The deployment of Fieldbus systems led to a significant reduction of cables and subsequently costs, as point-to-point connections are multiplexed over a shared communication medium [68]. However, the initial intention of introducing a single standard, ensuring generalized and compatible solutions failed as, over the years, several vendor-dependent standards have emerged. Prominent examples include PROFIBUS [70], INTERBUS [51] and Foundation Fieldbus H1 [72]. Most Fieldbus systems build upon the Ethernet standard, i.e., IEEE 802.3 [34], by modifying and extending it according to the special needs of industrial automation. Thereby, the most important modification concerns the Medium Access Control (MAC) protocol given that IEEE 802.3 implements Carrier Sense Multiple Access with Collision Detection (CSMA/CD), which is not suited for safety applications and critical closed-loop control, as the access to the medium is granted in a non-deterministic fashion.

Besides wired communication technology, the most recent focus of the industry lies on wireless communications, promising high flexibility in deployment and operation, while reducing installation and maintenance costs [1]. Although wireless communications, e.g., WiFi or Bluetooth, is the prevalent communication technology in home and office environments, it has not been widely adopted for industrial automation so far, because of the missing communication guarantees [75]. Therefore, wireless solutions specifically developed for industrial automation, such as WirelessHART [31] and ISA100.11a [36], aim to address the current needs of the industry by providing a comprehensive communication architecture comparable to the Fieldbus systems [50]. Such standards are either based on Bluetooth [32] or IEEE 802.15.4 [35], which both support low-rate and low-power data transmissions, including mechanisms to reduce latency and increase reliability.

An orthogonal approach to the IEEE-based standards, which mainly operate in unlicensed frequency bands, is the use of cellular standards such as 3GPP-LTE [4]. An advantage when relying on cellular networks is the already existing, globally available network infrastructure, which inherently supports mobility and security. Moreover, the operation in licensed frequency bands facilitates the coordination of coexisting devices. Nevertheless, currently deployed LTE standards do not account for reliable, low-latency M2M communications [54], therefore, a specific design requirement for the upcoming 5G standard is the support of the growing number of M2M applications. It is, however, not clear yet if the upcoming 5G standard will adequately address the requirements of the IIoT and ultimately be adopted in the domain of industrial automation.

In this chapter, we mainly focus on existing open standards, such as IEEE 802.15.4, to illustrate how communication in the IIoT could be realized. The requirements range from the provision of communication guarantees for local automation cells to the management and the communication of such local cells on the factory level, as well as via the Internet. Furthermore, this includes a worldwide interoperability of the presented solutions, comparable to the IEEE 802.11 standards in combination with TCP/IP for home and business environments.

2.3 Challenges for the Industrial Internet of Things

We identify different components that constitute a large-scale realization of communication within the IIoT. Each component introduces its specific challenges that must be addressed to achieve a coherent and flexible solution. In the following, we provide a concise overview of the different communication components of the IIoT and the challenges they introduce.

Local Automation Cells are areas within a factory hall where sensors, actuators, and controllers operate. These devices are interconnected to each other via a communication system, which should mostly rely on wireless links to account for flexibility and reduced deployment/maintenance costs. Furthermore, safety applications and critical closed loop-control (cf. Table 1) impose stringent communication guarantees, which should be mainly addressed on the physical and on the data link layer. Due to the broadcast nature of the wireless medium, devices within the same automation cell are, in most cases, able to directly communicate to each other.

Factory Halls consist of several local automation cells, where each cell has specific tasks and responsibilities. On the factory hall level, these cells must be interconnected to achieve a coordination and adaptability of the automation processes within the factory. This requires a device to have a unique address (at least on the factory hall level); additionally the communication system must provide routing to connect the devices to a plant controller and to allow for information exchange between devices in distinct cells. Furthermore, one or multiple gateways within a factory may provide accessibility to the Internet and thus allow for inter-factory connections. The supported application classes on the factory hall level and beyond usually include Class 2 and above (cf. Table 1).

Application Layer Communication enables access to IIoT services, e.g., to synchronize production processes across factories and further to adapt processes via Cloud services [78]. The main challenges in this context are similar to the ones in the traditional Internet of Things, e.g., efficiently connecting constrained devices to the Internet with standardized solutions.

In the following sections, we provide more details on the different architectural levels regarding the respective challenges on each level and in particular how current technologies could be applied and extended to solve these challenges.

3 Communication Within a Local Automation Cell

A local automation cell is an area within the factory hall in which all connected devices, i.e., sensors, actuators, and controllers, can communicate with each other via a local communication system. Devices that directly depend on each other, e.g., light barriers and safety switches, are typically covered by the same local automation cell. Within a local automation cell, we envision that several devices are wirelessly connected to each other via an Access Point (AP). The AP manages the network and simultaneously acts as a gateway to the factory network. Note that the communication within an automation cell may also be organized in a distributed way as mentioned later in this section.

The dominant communication technology on the automation cell level should be wireless, enabling a high flexibility for moving machine parts while reducing costs due to fewer cables that must be deployed and maintained, as already mentioned in Sect. 2.2. The main challenge for the local automation cell is thus to support application Classes 0 and 1 (cf. Table 1) using inherently faulty wireless links. Mechanisms that increase reliability and reduce latency are primarily implemented on the physical and the MAC layer which allows tailoring them to the hardware and to the characteristics of the communication medium. In the remainder of this section, we present different techniques that are applied on the physical and the MAC layer to achieve QoS in the context of industrial automation and provide examples of actual implementation of these schemes.

3.1 Diversity Schemes

In general, the propagation of a wireless signal is deteriorated by fading, path loss, shadowing, and interference, their magnitudes strongly depend on the environment. It is known that factory halls have unfavorable propagation characteristics for radio signals [75]. One possibility to mitigate some of these negative effects—fading and interference—is to introduce a fading margin, i.e., to increase the transmit power. However, this approach is not energy efficient, especially when considering the growing number of battery-operated devices. Moreover, this approach increases the co-channel interference for neighboring systems and is thus not suited for a large scale wireless coverage [57]. A well-researched alternative mitigating fading and interference is to use diversity schemes, i.e., leveraging redundant uncorrelated transmissions of a signal to exclude those transmissions that are currently in deep fading [53]. Diversity may be applied in the frequency, time, and space domain and possibly be combined to further increase the transmission reliability. In the following, we provide a short overview of the respective diversity schemes and relate their application to IIoT scenarios.

3.1.1 Frequency Diversity

To exploit frequency diversity, the transmission reliability is increased by using different uncorrelated frequencies. One example is channel hopping, where, during a transmission, the sender repeatedly changes the channel using a predefined hopping sequence. Thus, the time that a transmission is disturbed by narrow band interference is reduced, as the sender only remains on a given channel for a short time. Additionally, the sender may apply blacklisting to exclude channels with poor link qualities from the channel hopping. On the downside, channel hopping requires a precise synchronization between sender and receiver as both must know the current hopping sequence. Both WirelessHART and ISA100.11a apply channel hopping in combination with channel blacklisting to mitigate interference.

In the context of the IIoT, the main advantages of frequency diversity lie in the efficient implementations that are available for such schemes [73]. This enables even small devices with limited hardware in terms of CPU, memory, and transmission antennas to make use of frequency diversity and thereby increase the transmission reliability. However, when deploying multiple networks using uncoordinated channel hopping in close proximity, the probability of interference increases. Alternatively, the frequencies of neighboring networks could be synchronized manually, which, however, impedes dynamic deployment of cells and scales poorly with various of such networks.

3.1.2 Time Diversity

The basic idea of time diversity is to add redundant transmission symbols to increase the probability of successfully decoding a message, even when some symbols are not correctly received. A known technique for time diversity is Forward Error Correction (FEC), which can be further differentiated into block codes and convolution codes [27]. While block codes are applied on fixed sized blocks of data, convolution codes operate on bit streams and thus provide a higher flexibility. Additionally, to mitigate the negative effects of error bursts, interleaving may be applied. Successive transmission symbols are scrambled over time such that two consecutive symbols are not affected by the same error burst.

The use of FEC is indispensable to achieve reliable communication in wireless transmission systems. However, it must be noted that time diversity is in direct conflict with low latency guarantees for industrial automation, particularly in the presence of slow-fading channels, as the channel code length and thus the transmission time must be increased. Time diversity should hence be applied in combination with other diversity techniques to leverage the right trade-off between the reliability benefit and the time costs. An example of such a combination is the implementation of space-time codes, which in addition to time diversity also exploit spatial diversity by using multiple antennas at the transceivers [67].

3.1.3 Spatial Diversity

To benefit from spatial diversity, a transmitter uses spatially uncorrelated transmission paths to convey a message to a receiver [15]. This is typically achieved by including multiple antennas at the transceivers. A prominent example in this context is Multiple-Input and Multiple-Output (MIMO), which leverages multipath propagation to combine different, independent data streams of the same message to increase the signal strength at the receiver. Foschini et al. [20] showed that MIMO significantly increases transmission reliability depending on the number of used antennas. Therefore, in recent years, MIMO has been adopted by several standards, including current IEEE 802.11 extensions as well as Long Term Evolution (LTE).

A special type of spatial diversity is cooperative diversity, where the stations form a virtual antenna array through cooperation [39]. When a sender transmits a message to a receiver, a third station overhears the message and retransmits it in case the receiver could not decode the original transmission. Hence, the retransmission is performed via an additional, uncorrelated transmission path. This is particularly beneficial for retransmissions within a short time frame, as a retransmission via the same link would result in a similar error pattern due to the coherence time of wireless channels [43]. Cooperative diversity thus exploits the very nature of omnidirectional wireless transmissions: Every station within the same broadcast domain may overhear ongoing transmissions from other stations and thus act as a relay if necessary.

For the IIoT, it is especially beneficial that cooperative schemes are suited for devices with limited hardware features, i.e., with only one antenna, as the IIoT strives for wirelessly connecting a large number of diverse devices. Many of these devices, e.g., distributed sensors, may be constrained in space and production costs, which preclude adding additional antennas to achieve spatial diversity. Moreover, it has been shown that cooperation can be leveraged to achieve high reliability within a low latency bound in an entirely distributed setting [17], as well as in a setting with centralized relaying [58]. Additionally, *multi-user* MIMO can be implemented at multi-antenna APs to serve several (single-antenna) stations simultaneously [64].

3.2 Medium Access Control

In the case of a shared communication medium for the local automation cell, the exclusive access of a station to the medium must be coordinated between all participating stations. In general, a Medium Access Control (MAC) protocol is either organized in a centralized or in a decentralized way [24]. Moreover, MAC protocols can be further differentiated into contention-based access protocols, e.g., Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA), and schedule-based access protocols, e.g., Time Division Multiple Access (TDMA). A combination of both, sometimes referred to as hybrid, is also possible. The performance of a MAC protocol strongly depends on the considered application scenario and thus on the performance metrics relevant for the considered scenario. In the following, we present typical performance metrics of MAC protocols, explaining for each their impact in the realm of IIoT applications. Afterward, in Sect. 3.2.2, we shortly introduce the MAC protocol of IEEE 802.15.4, which is one of the de-facto standards for wireless sensor networks [5].

3.2.1 Performance Metrics with Regard to the IIoT

Application scenarios within the IIoT have very different requirements that must be met by the communication system, ranging from configuration updates and monitoring tasks to safety applications, where the latter is challenging to achieve with state-of-the-art MAC protocols [66]. Ideally, a MAC protocol for a local automation cell is able to fulfill all different requirements, depending on the current needs of the application layer. The most important performance metrics for MAC protocols in the IIoT are as follows.

Reliability refers to the fraction of transmitted data that correctly reaches its destination. In closed-loop control, reliability is an important optimization goal, as such control loops are neither error- nor loss-tolerant.

Delay indicates the average time a packet awaits transmission in the MAC queue. It thus strongly depends on the average time for a station, which wants to transmit data, to get access to the channel. Besides reducing the delay, an important optimization goal in mission-critical scenarios is to minimize the variation in the delay (jitter), in order to provide latency guarantees.

Throughput describes the fraction of the channel capacity that is used for data traffic. When maximizing the throughput, the amount of MAC management traffic and idle channel times must be reduced. For most local automation cell applications, however, the throughput is not the primary optimization goal.

Power Consumption is particularly important for battery-operated devices. An energy-efficient MAC protocol can significantly reduce the power consumption of single devices, e.g., by introducing inactive periods in which the stations turn off their transceivers. However, optimizing power consumption often occurs at the expense of throughput and delay.

Fairness is achieved when each station may access the channel under the same conditions. However, for the IIoT, different traffic priority classes may be introduced, i.e., safety-relevant traffic is more important than monitoring traffic. When introducing priority classes, a fairness metric typically only applies to the traffic of the same priority.

3.2.2 Example: IEEE 802.15.4

IEEE 802.15.4 supports a centralized star topology as well as a decentralized peer-to-peer topology [35]. Depending on the topology, the MAC protocol of IEEE 802.15.4 is organized either in a centralized or in a decentralized way, where the latter is essentially a CSMA/CA protocol. In the centralized mode, a Personal Area Network (PAN) coordinator is responsible for the MAC protocol, which is organized through superframes. To signal the beginning of a superframe, the PAN coordinator periodically sends a beacon, which also serves as a synchronization reference for the associated stations. A superframe is then further divided into a Contention Access Period (CAP) and a Contention Free Period (CFP). Figure 2 shows the structure of a superframe as defined in the IEEE 802.15.4 standard. The time between two superframes is referred to as inactive period, as the stations may turn off their transceivers to save energy. It has been shown that a significant amount of energy can be saved by increasing the inactive period. However, this comes at the price of a higher latency and lower throughput [42].

The active portion of a superframe comprises exactly 16 equal-length time slots which are split between the CAP and the CFP. In the CAP, the stations compete for the time slots using CSMA/CA, therefore, the access to the shared medium is not deterministic. The CFP, in contrast, supports time-critical data traffic with Guaranteed Time Slots (GTSs). A GTS consists of one or more time slots which are grouped together and centrally assigned on demand to a station by the PAN coordinator. When the PAN coordinator receives a request for a GTS from a station, it checks whether it has enough transmission resources and, if so, allocates the

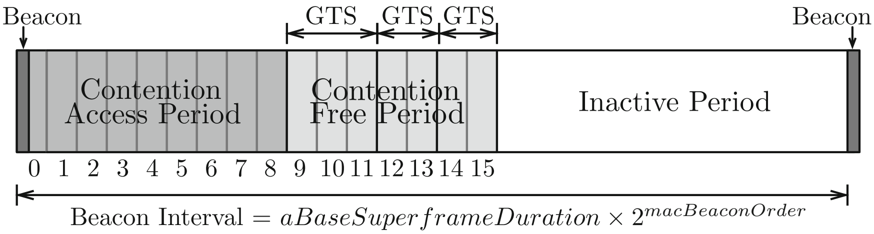


Fig. 2 Beacon interval of IEEE 802.15.4 including structure of superframe. The superframe consists of a beacon, a Contention Access Period (CAP) and a Contention Free Period (CFP)

requested time slots as a GTS for the station. The current assignment of GTSs is propagated from the PAN coordinator to the stations through the beacon message. A station may only transmit during a GTS if the allocation is confirmed in the beacon message of the current superframe.

WirelessHART, which builds upon IEEE 802.15.4, further adapts the MAC protocol to be able to provide stringent communication guarantees [63]. The protocol exclusively uses GTSs with a fixed length of 10 ms. The assignment of time slots is centrally performed to ensure deterministic and time guaranteed access to the medium. To increase reliability, WirelessHART uses channel hopping in combination with channel blacklisting and Automatic Repeat ReQuest (ARQ).

In comparison to WirelessHART, ISA100.11a, which is also based on IEEE 802.15.4, implements GTSs for time-critical traffic as well as CSMA/CA with priorities for retransmissions and maintenance traffic [16]. ISA100.11a defines flexible slot lengths between 10 and 12 ms and, furthermore, allows for devices to have multiple superframes with varying lengths depending on their traffic load. Moreover, a superframe may contain a CAP, where the CSMA/CA mechanism was adapted to support different priority classes. The priority classes are controlled by introducing individual delays, depending on the priority, before a station is allowed to perform a clear channel assessment.

4 Communication Within the Factory Hall and Beyond

This section addresses higher communication levels in the context of establishing communication within a factory hall, between multiple production cells and beyond. Possible scenarios for this type of communication are either the interconnection of multiple production cells via wireless links or a wired backbone, realized with multi-homed gateways, as depicted in Fig. 3. As soon as the communication between devices exceeds the distance of one communication hop, mechanisms for *routing* and *global addressing* need to be deployed. In this section, we focus on mechanisms that tackle these tasks on top of IEEE 802.15.4 [35] networks, namely WirelessHART [31], ISA100.11a [36] and 6LoWPAN/RPL [44, 76].

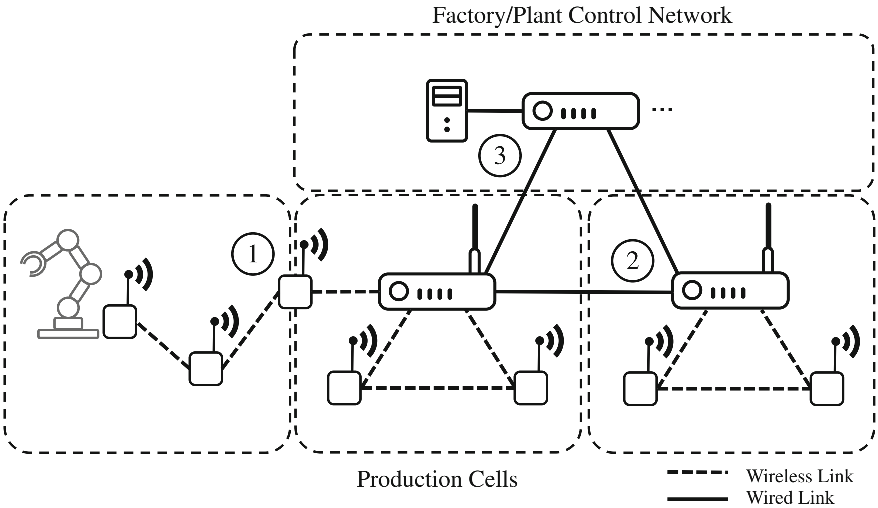


Fig. 3 Communication scenarios between production cells within the factory or across the whole factory site. Individual cells are either (*1*) connected through wireless links or (*2*) by multi-homed gateways, connected via a wired link. Ultimately, these are connected to (*3*) the control and automation network of the factory

4.1 Routing and Addressing Background

Before focusing on the available techniques in the aforementioned mechanisms, we provide a brief overview of routing in general. Depending on the topology dynamics of a network, different routing schemes can be applied. In a fixed wired setup, a static scheme can be deployed. Each node maintains a routing table with an entry for each possible destination and how to reach it. However, in the envisioned IIoT the production setups should be adaptable in a flexible manner, supported by the use of wireless communication links. Therefore, a certain degree of adaptivity of the routes themselves is required, rendering a static solution infeasible. Discovering routes between nodes without a static configuration can either be done infrastructure-assisted or ad-hoc. In an infrastructure-assisted solution, a central instance collects information from all nodes in the network, such as discovered neighbors, connection quality or traffic demands. Based on the aggregated global information and a set of metrics, the central instance is able to compute the optimal path for each individual source-destination pair in the network, which it then distributes in form of a routing table to the devices. Examples for these metrics are the number of hops towards a destination, available link throughput, latency, or, especially for battery-powered devices, energy. The advantages of this solution are the availability of global information and the mitigation of required computation at the nodes. Moreover, because the nodes send updates of the aforementioned local information, this solution provides a certain degree of adaptivity. However, the disadvantages are, depending on the setup, a non-negligible computation overhead at the central entity and a single point of failure, if no backup or fallback mechanism is provided.

In an ad-hoc network, the participating nodes form a self-managing network, i.e., each node may have to route and forward data, such as in Mobile Ad-Hoc Networks (MANET) [6] or Mesh networks [2]. The used routing protocols in such networks can be divided into *proactive* and *reactive* protocols. Although hybrid solutions exist, we focus on these basic designs. In a proactive protocol [11], the goal is to maintain an overview of the entire network at each point in time. To this end, nodes periodically broadcast their own routing tables, based on local neighborhood information obtained from periodic discoveries. The advantage of this approach is the availability of routing information at all times. However, the major disadvantages are the slow convergence in the presence of restructuring or link failures, as well as the overhead of control and maintenance traffic, especially when the actual amount of data, such as sensor information or actuator control messages, is sparse. If a routing approach uses well-maintained routing tables for the whole network, such as in proactive protocols, the decision of how to reach the destination is *hop-by-hop*, i.e., only the destination address needs to be contained in the packet itself.

A reactive approach [49] starts flooding the network with *route requests* on-demand when a route is needed. Intermediate nodes that intercept these requests will add information including their own address to the packet before forwarding it further, thereby recording the path. Once the destination is reached, typical reactive approaches send back a route reply via the reverse route that was received in the route request at the destination. An example is shown in Fig. 4, where Node A requests a route to D. Intermediate nodes receive this request and further broadcast the message, until it reaches the destination.

Advantages of reactive approaches are that topology information is only sent when needed, thereby reducing constant communication overhead and saving local memory on the devices. On the down side, this approach introduces additional latency because routes are only calculated on-demand. As the data source includes the whole route into the packet after discovery, this mechanism is called *source* routing.

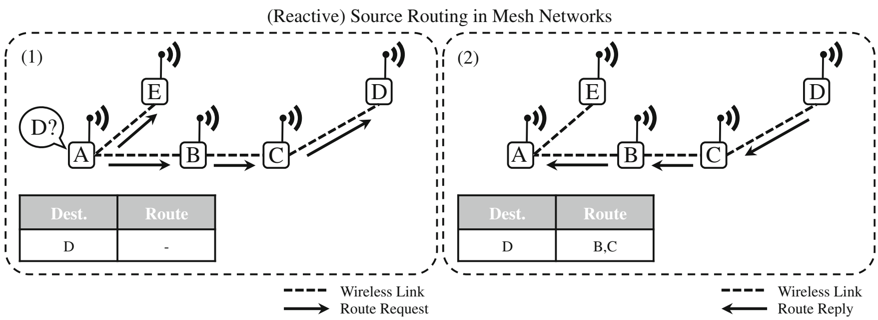


Fig. 4 Route discovery in reactive mesh routing. Node *A* requests a route to *D*, which is discovered and after the route reply from *D* (temporary) stored in *A*'s routing table

4.2 Challenges for Routing and Addressing in the IIoT

Networks that are created from a set of constrained and wireless devices are typically called Low-power and Lossy Networks (LLN) [52]. These networks face many requirements which we discuss in the following.

Traffic Demands. Depending on the application or traffic class (cf. Table 1), various levels of reliability in terms of availability and latency are necessary. A non-critical application may not be harmed by packet loss or delayed packets, while a machine control system might require nearly loss-free transmissions as well as low latency. To this end, a routing protocol must be able to compute paths on various metrics, e.g., throughput, latency or stability, to accommodate the requirements of various applications.

Deployment, Performance and Management. Because the number of IIoT devices used in an envisioned factory setting might be very large, a manual configuration, especially with respect to the desired possibility to allow for dynamic production setups, is infeasible. To this end, the routing protocol should not require any pre-configuration other than information about the device itself, e.g., a device identifier or allowed radio channel, but rather be capable of fully discovering its surroundings, such as available neighbors or gateways, and leverage this information for an automatic setup.

Device-Awareness. As soon as packets need to be forwarded over multiple hops, intermediate devices take part in the routing process. In the envisioned IIoT, these are constrained devices with possible limitations in terms of power supply, memory and computation. Therefore, a routing protocol must incorporate this information in its routing decisions, e.g., compute multiple routes and offer alternatives to limit the number of forwarding steps of individual nodes, thus achieving load-balancing.

Adaptivity. Dynamic topologies, caused by the mobility of machines and workpieces that are equipped with IIoT devices, and the possibly harsh environment of factory halls, have an effect on the link quality and on connectivity. Therefore, it cannot be generally assumed that a route is always available in such a wireless setup.

Message and Storage Overhead. Depending on the setup, additional maintenance messages may introduce a significant overhead. If nodes in the network only sparsely send data, the aforementioned maintenance traffic might outnumber the actual application traffic. This additional communication overhead introduces a non-negligible amount of power consumption which has to be taken into account. In addition, memory constraints on the devices might not always allow to store routing tables of a certain size.

4.2.1 IPv6 and 6LoWPAN

Up to now, we only focused on routing itself. However, in order to reach a certain destination, we need an identifier, i.e., an address. In the predecessor of the IoT, so-called wireless sensor networks (WSNs), local addresses were used that were only valid within a certain cell, thus not allowing a convenient global addressing scheme. To achieve such a global scheme and to support the integration of these networks into larger networks, i.e., a company-wide network or even the Internet, the IETF proposed to use already established standards for addressing, based on the Internet Protocol (IP), with the current version being IPv6 [14]. Such standardized solutions, instead of closed and proprietary solutions, make it possible to use well-maintained and widely-adopted mechanisms. In addition, going beyond the communication within a factory is a key enabler for the IIoT, enabling the integration into existing network structures. However, there are a few obstacles, raised by the used constrained devices and the incorporated physical and data link layer mechanisms, that hinder the direct use of IPv6, which we discuss in the following.

Addresses used in IPv6 have a length of 128 bit, and are constructed in a hierarchical way. Typically, it consists of a 64 bit prefix, that identifies the respective sub-network and a 64 bit interface identifier that has to be unique for the aforementioned sub-network. This interface identifier is formed from the already available physical interface address, i.e., the MAC address. The IPv6 standard requires the Maximum Transmission Unit (MTU) in an IPv6 network to be at least 1280 Byte. Compared with the available maximum of 127 Byte in IEEE 802.15.4 networks, this already induces the need for an additional adaptation for the use of IPv6 [45]. In addition, depending on the setup, a non-negligible amount of these 127 Byte is occupied by the protocol itself. In the worst case, i.e., considering the maximum frame overhead of IEEE 802.15.4 (25 Byte), the possible maximum link layer security (21 Byte), a full sized IPv6 header (40 Byte) and using a lightweight transport protocol, such as the User Datagram Protocol (UDP) (8 Byte), only 33 Byte are left for actual application data. To be able to use IPv6 in this highly constrained setting, the IETF initiated a working group that focused on how to realize IPv6 over Low-power Wireless (Personal) Area Network (6LoWPAN) [38, 44].

The standard is realized as the 6LoWPAN adaptation layer that is located between the data link and the network layer of traditional communication stacks. Implementations often realize IPv6 and the 6LoWPAN adaptation as single layer, thereby, softening this separation [60]. One major task of this layer is the fragmentation and re-assembly of packets, to virtually achieve the desired MTU of at least 1280 Byte for IPv6. Another important technique is header compression [7, 28], which is capable of reducing the protocol overhead significantly. These techniques allow saving valuable space for application data itself, such as control or monitoring information. However, the amount of saved space depends on the scenario, e.g., if two devices in the same network want to communicate, the network prefix is known and can be omitted. In contrast, addressing beyond network borders does not allow to remove certain prefixes.

4.3 Addressing and Routing in Standardized Protocol Stacks

Now, we focus on the design and implementation of routing in available protocol stacks and the typical network structures envisioned in these approaches. In WirelessHART, the routing functionality resides on the network layer, similar to traditional protocol stacks. WirelessHART follows an assisted approach, using a so-called *network manager*, which is responsible for creating routing information for all devices. To this end, this network manager has to collect the necessary information from all devices in the respective sub-network [48]. A typical topology consists of several wireless devices, forming a multi-hop setup. The typical overall setup is shown in Fig. 5a. As soon as a device joins a WirelessHART network, the network manager assigns a 16 bit (2 Byte) network address. This is mainly done to save the overhead of sending the full physical address of the device in each packet, which is typically 8 Byte for WirelessHART. In addition, the joining device sends a list of discovered neighbors. The network manager might be an additional entity or co-located with a gateway that connects this sub-network to other sub-networks or to the factory's backbone. To allow for routing within this sub-network, WirelessHART offers multiple schemes. The simplest case is *source routing*, where the complete route is included in the data packet. This is typically used by the network manager for diagnostics.

In order to allow the devices to communicate in the sub-network, *graph based routing* is used. Therefore, the network manager constructs several directed graphs for the respective sub-network. Graphs are not unique and may overlap, which is a desired feature, e.g., to have alternative routes. For more information about the

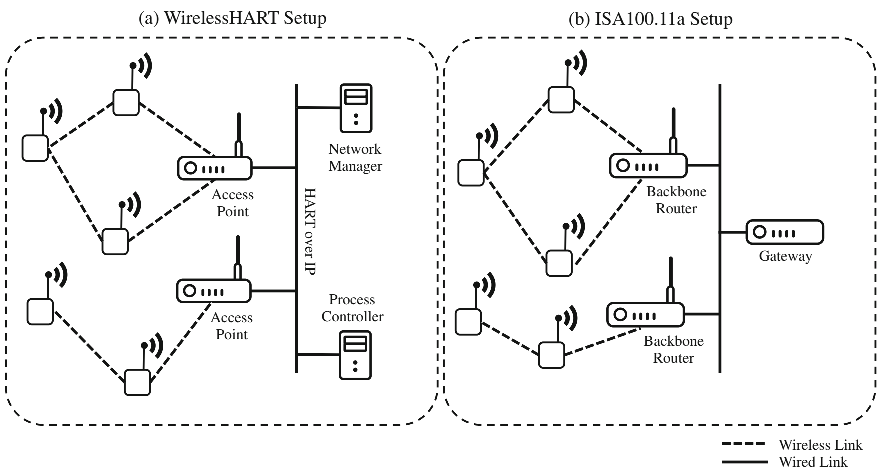


Fig. 5 Exemplary WirelessHART and ISA100.11a setups, including wireless devices and the necessary management entities. Adapted from [47]

actual algorithm for the graph construction, we refer to [25]. Please note that the WirelessHART standard [31] itself does not define any specific algorithm for the construction of these graphs. Once these graphs are constructed, they are deployed on the devices. As a device might be part of multiple graphs, the devices obtain an additional *graph ID* for each graph. Moreover, the WirelessHART standard defines a network as configured properly, if all devices have at least two devices in the graph to which they may send packets, called associated neighbors, which ensures redundancy and enhanced reliability. If a device wants to send data to another device, it selects the according graph ID. This graph ID is then also written into the network packet. On intermediate devices, the packet is forwarded based on this graph ID until it reaches the destination, as shown in Fig. 6.

Recent work has investigated the performance of source routing and graph routing with a set of empirical test cases [59]. The results indicate that graph routing outperforms source routing in the respective setups in terms of worst-case reliability, at the cost of a higher energy consumption and an increased latency.

Notably, WirelessHART is designed as a wireless extension for HART to send HART commands, e.g., read or write dynamic variables or calibrate. Thus, the application layer targeted to be served by the underlying layers is already specified, which is a major benefit with regard to interoperability with already deployed HART solutions. To integrate a WirelessHART network into larger network structures based on IP communication, specialized gateways that translate between the two different standards need to be deployed.

ISA100.11a [36] follows a similar setup as WirelessHART (cf. Fig. 5b). Wireless devices form a multi-hop network, called Data Link (DL) subnet, which is connected via a *backbone router* to a backbone link that connects multiple cells. In addition, this backbone link is connected via an *ISA100 Gateway* to the factory network. A central entity, called *Network System Manager* (NSM) is also attached to the backbone and takes care of creating the routing graphs for a respective sub-network. Although there also exist multiple graphs for the network as in WirelessHART, here these are used to send different types of traffic. For example,

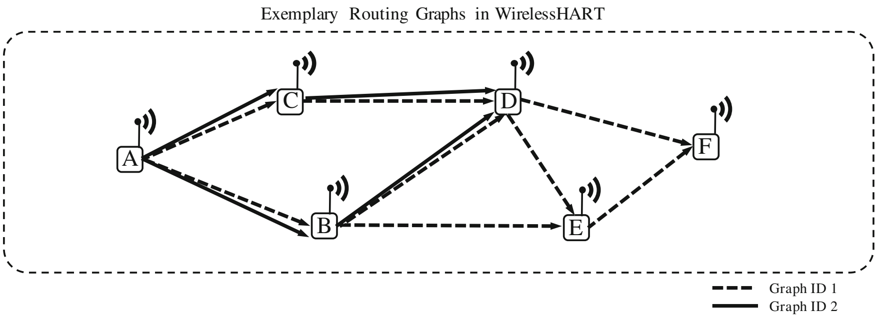


Fig. 6 A communicates with *F* using the graph with *ID 1* (*dotted line*). To send a packet on that graph, *A* can either forward to *B* or *C*. Ultimately, following this graph will end at *F*. Similarly, in order to send data to *D*, *A* can use the graph 2 (*solid line*)

event data, such as aperiodic reports, may require sending bursty traffic, whereas *periodic data*, e.g., sensor readings, may require timely delivery. To create these graphs, the NSM collects information about the wireless neighborhood from the clients and also incorporates requirements provided by the designer of the network, such as bounds for latency or minimal throughput. The resulting graphs are marked with a *contract ID*, which can then be used to indicate the requirements for a particular transmission. To increase the reliability, the NSM also generates backup graphs and may update the graphs based on metrics, such as link quality or battery level. Additionally, ISA100.11a also incorporates source routing.

Notably, routing within a DL subnet takes place on the data link layer. To this end, each device in the DL subnet is assigned a 16 bit DL subnet address for the purpose of *local* addressing by the NSM. For routing beyond the backbone router, where the DL subnet is terminated, the network layer (NL) is used. The NL in ISA100.11a can use addresses of a length of up to 128 bit, such as IPv6. To this end, the network layer in ISA100.11a uses the 6LoWPAN adaptation layer. To achieve end-to-end connectivity beyond a local link, backbone routers that are 6LoWPAN-enabled have to be deployed, which re-assemble the packets and translate them to the normal IPv6 format for further routing and processing. For the routing inside an ISA100.11a mesh, routing is based on the aforementioned graphs and the information in the data link layer. Therefore, IPv6 is not used in the wireless network directly. To distinguish between routing mechanisms on different layers, mechanisms on the data link layer (Layer 2) are typically called *mesh under* and on the network layer (Layer 3) are called *route over*. Although ISA100.11a and WirelessHART share some similarities, the overall goal of ISA100.11a is to achieve more flexibility. To this end, the application layer does not specify a protocol for automation and control purposes, such as in WirelessHART. It rather defines how to provide interfaces to such protocols, enabling the realization of own protocols or to realize a tunneling of existing protocols [47].

In order to establish a protocol for Routing Over Low-power and Lossy links the IETF formed the ROLL working group, which specified the Routing Protocol for Low-power and Lossy Networks (RPL) in [76]. RPL is a so-called *distance vector* routing protocol for IPv6 in LLNs and uses Destination Oriented Directed Acyclic Graphs (DODAG). These graphs can be built to optimize a certain goal, e.g., paths with best expected transmissions or avoiding battery-operated nodes. The root of such a graph should be located at a LoWPAN Border Router (LBR), connecting this sub-network to a larger network. To connect cells to a larger network, at least one LBR per cell is deployed. The graph building process typically starts at these LBRs and uses a set of management messages. After the root started advertising its presence, nodes in communication range will further process these messages based on their configuration. This configuration may contain information about certain objectives for the graph or path costs. If a node decides to join the graph, it has a route *upwards* to its parent, which is the graph's root in this initial case. Moreover, the node computes a *rank* in this graph that increases towards the leafs, used to indicate the hierarchy inside the graph. Depending on the configuration, i.e., acting as a router or a leaf, it will either announce the graph further or not. This process continues, resulting in a graph as depicted in Fig. 7.

Up to now, each node can reach the root of the graph by simply sending a message *upwards* to its parent node, thereby achieving *Multipoint-to-Point* communication. To achieve a traffic flow from the root or an intermediate node towards a leaf, a message is sent *downwards*. Therefore, additional management messages are used that carry information about the reachability of nodes. They include a prefix and a freshness indication and are sent by the individual parents upwards the graph. Intermediate parents on the path to the destination may also aggregate this information or remove redundant messages. Once this information is complete, and in combination with the upward routing, *Point-to-Point* routing in the graph is possible.

However, the efficiency of the routing depends on the amount of information that intermediate devices store permanently. RPL offers intermediate devices to work in two different modes, a *storing* and a *non-storing* mode. In the storing mode, a node keeps routing information, e.g., about the reachable children. In the non-storing mode, intermediate nodes do not store any routing information at all, except knowledge about their parent node. Assuming a node A wants to send data to a node B and the intermediate nodes use a storing mode. If B is not directly reachable, A sends the data upwards the graph to a so called *common ancestor parent* [71], which, in turn, will then forward the message down to B. In the non-storing mode, the message of A has to be forwarded to the root of the graph and then the root uses a form of source routing to reach B.

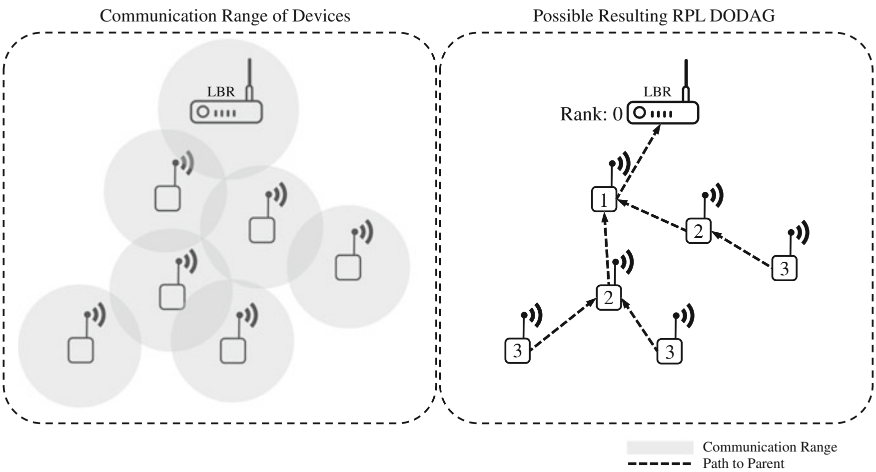


Fig. 7 Exemplary scenario for the formation of a Destination Oriented Directed Acyclic Graphs (DODAG) for routing with RPL. The LoWPAN border router initializes the creation of a DODAG, also called RPL instance, by announcing itself as root, i.e., node with *rank 0*. Depending on the configuration, some nodes will forward this information with themselves as roots (parents) for the respective sub-DODAG. The rank of the nodes must monotonically decrease towards the root and can be seen as the location of a node within a DODAG

The ability to participate in a non-storing mode is a desirable feature for the use of highly constrained devices. Nevertheless, the additional routing steps in the non-storing mode add a certain amount of latency. In RPL, a mixture of nodes either supporting storing or non-storing is possible. Similar to ISA100.11a, a multitude of graphs for different purposes can exist, each one called an *RPL Instance*.

In addition, RPL defines *local* and *global* repair mechanisms. If a link fails and a node has no alternative route towards the root, a local repair process is performed to find an alternative. This might ultimately lead to a sub-optimal graph, depending on the overall objective, as the node might ignore certain metrics but rather aim for connectivity again. To cope with the problem of a diverging graph, the root can trigger a global repair, re-starting the graph building process.

Addressing in networks using a combination of 6LoWPAN/RPL is done via IPv6. RPL itself offers neighbor discovery mechanisms and leverages auto-configuration features of IPv6 [62, 69], thereby facilitating the process of initial configuration. Due to the use of the adaptation layer, the LBR has to reassemble fragmented packets if they have to be routed beyond the borders of local cells, e.g., to a service located in the Internet.

5 Application Layer Communication

IP-based routing and 6LoWPAN are the foundations that are used to interconnect factories. Having a globally unique addressing allows identifying every single node in the IIoT. An edge router on the local factory network translates 6LoWPAN to fully fledged IPv6 packets and vice versa, thus, no application layer gateways are required.

Still, to allow meaningful data exchanges between devices and services, additions on top of IP are required. To this end, several application layer protocols have been proposed, e.g., CoAP, MQTT(-SN), XMPP, AMQP, DSS, and HTTP [3]. Some of these protocols also include routing as already provided by RPL/6LoWPAN, others, often motivated by traditional web protocols, are unsuited in constrained environments. Nevertheless, all of these protocols offer a framework that enables application developers to express and communicate data in a standardized fashion while offering a comprehensive communication model.

Within IoT applications, CoAP and MQTT(-SN) have proven to be the most prominent data exchange protocols. The Message Queuing Telemetry Transport for Sensor Networks (MQTT-SN) was originally designed for Wireless Sensor Networks (WSNs) to enable a publish/subscribe system [30]. It especially targets low-end battery powered devices that have only a limited bandwidth available. In contrast to its origin MQTT, which relies on TCP as its underlying transport, MQTT-SN does not require TCP and even works independently of an IP network. Thus, it is well suited in situations where it is impossible to deploy IP or 6LoWPAN. However, MQTT and also MQTT-SN require the presence of a so-called broker that acts as a server to which the clients (sensors/actuators) publish their data or subscribe to. In the case of MQTT-SN, even a gateway is required that translates between MQTT-SN and MQTT in-between the broker and client [65].

However, in the vision of an all IP IIoT, MQTT-SN, with its requirements for additional components, and MQTT, requiring TCP, may not be the first choice. It has been shown that a TCP stack and TCP's protocol behavior pose too high demands on extremely constrained devices and thus are not a feasible choice [12]. To grasp even the most constrained devices, the next section focuses on the *Constrained Application Protocol (CoAP)* [61] as specified by the IETF Constrained RESTful Environments (CoRE) working group. Afterwards, we highlight cloud technology as an enabling mechanism that allows to gather and structure large amounts of sensor data while simultaneously offering control over the entities within the IIoT.

5.1 The Constrained Application Protocol

CoAP offers similar semantics to the user as the Hypertext Transport Protocol (HTTP) [19] but in a much more lightweight fashion [79]. Therefore, it operates on top of resources identified via Unified Resource Identifiers (URIs), e.g., `coap://host/resource`. Thus, similar to how different resources are located in the World Wide Web (WWW), different resources can be located on a single server using different paths. Moreover, the strict client/server model used in HTTP is softened, in M2M applications devices may operate as both, client and server.

CoAP should be operated on resource constrained devices, therefore, CoAP encourages to follow a REST-style [18] operation of services. REST allows servers to be stateless in their operation (saving costly resources): either context is implied through the resources accessed or all necessary information is included in a request.

In contrast to HTTP, CoAP offers a simply-parseable and tightly represented header format that has less overhead [37]. The header, in its minimal configuration, has only 4 Byte, however, it is extensible by adding binary options. Colitti et al. [12] report that typical header sizes, including extensions, range from 10 to 20 Byte.

Still, CoAP is not a way of compressing HTTP [61], it is fundamentally different from HTTP in the sense that its design assumes the use of the User Datagram Protocol (UDP) [55]. Thus, compared to the Transmission Control Protocol (TCP) [56] (as used with HTTP), it is inherently unreliable. To overcome these shortcomings, CoAP offers simple reliability mechanisms: two bits in the header indicate the desired transport mechanism of a message [61]. Messages that are flagged (in these two bits) *confirmable* (CON) need to be *acknowledged* (ACK), therefore, a CoAP message may contain a *Message-ID* to correlate signaling traffic. This Message-ID is also used to detect duplicated messages even for *non-confirmable* (NON) messages that offer no means of reliability. In case a confirmable message does not receive an acknowledgement within a certain time, a retransmission that is subject to an exponential backoff is initiated [12].

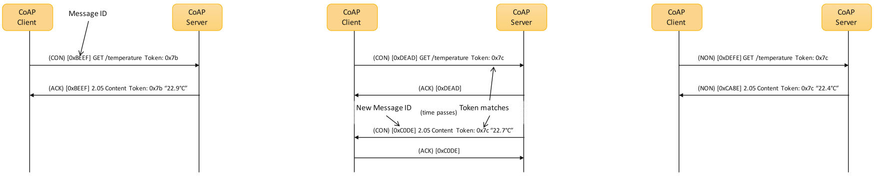


Fig. 8 *Left* A CoAP request that is directly answered by piggybacking the response on the ACK. *Middle* A delayed response to a request, please note the use of a fresh Message-ID and the same Token. *Right* A non-confirmable request and response, using different Message-IDs but the same Token. (cf. [61])

In addition to the Message-IDs, CoAP uses *Tokens* to identify and correlate requests and responses. So, CoAP, even though standardized as a single protocol, operates in a two-layered fashion using Message-IDs for correlation of signaling traffic and tokens to identify a semantic correlation of content. Figure 8 shows three different CoAP requests. On the left, a CoAP request is sent using a confirmable message, the server answers by acknowledging the receipt of the message, at the same it piggybacks the reply onto the acknowledgement. Here, the return code `2.05` is used to indicate that the message contains content, similar to a `200` OK in HTTP. If the resource was not available, the server would answer with a `4.04` indicating that the resource cannot be found. A second possibility for the CoAP server to answer is depicted in the middle: Instead of directly replying, the server may choose to delay an answer, e.g., because it cannot yet answer the request or is busy otherwise. In that case, the CoAP server acknowledges the receipt of the request, using the same Message-ID as contained in the request, and once it is capable of sending a reply, it sends a new confirmable message containing the response that can be correlated to the earlier reply using the Token. This message must, in turn, be acknowledged by the recipient. The right side depicts a non-confirmable exchange. Messages are not acknowledged, and each transfer includes a new Message-ID, thus again showing that the Token is used to correlated semantic content.

In the following sections, we explain how larger requests and replies may be sent, how changes of resources can be efficiently monitored and why CoAP can be seamlessly integrated with existing web and cloud infrastructures.

5.1.1 The Block Mode

Typically, CoAP messages carry only a few bytes of payload, e.g., sensor reading or simple instructions, thus even in the realm of IEEE 802.15.4, there is no need for fragmentation. However, sometimes it might be required to send larger chunks of data, e.g., for a firmware update, therefore, CoAP offers a Block mode [9]. Even though UDP allows for payloads of up to 64 kB, using these large payloads leads to

IP fragmentation or 6LoWPAN adaptation layer fragmentation [8]. As IP fragmentation requires the constrained devices to keep all packets in memory before they can be reassembled, this is not desirable, especially if the data is to be read as a stream, i.e., it is not required that all of the data is in memory at the same time. The block-wise transfer in CoAP is currently still only defined as an Internet Draft [8]. Effectively, two different block options, one for requests (typically using POST or PUT) and one for responses (e.g., when requested via GET) are defined. Both block options contain the block size (*sz*), whether more blocks are following (*m*), and the relative number of the block in a sequence (*nr*).

Figure 9 gives a simple example of the block mode. If a resource is to be transmitted in block mode, the server first acknowledges the receipt of the request but includes the block option, announcing the first block (nr = 0) and typically that more blocks follow (m = 1), it also suggests a block size (sz). The payload that is transmitted must always exactly match the block size (except for the last packet (m = 0)). After having received the first block, the client can inspect the m-flag to see if other blocks are available. It can then proceed to request the next block (here nr = 1). By including the same block size in that request, the client agrees to the block size proposed by the server. If the client is incapable of receiving the proposed block size or if it is more efficient for the client to use a smaller block size, it may request a smaller block size, effectively downgrading the block size for this transmission. The client may also anticipate the size of the response (e.g., due to context or link discovery), then the client may already include the block option in the initial request proposing a block size.

The block mode works analogously for requests that contain a large amount of data that is pushed to the server, e.g., using POST. Moreover, [8] specifies how the block mode may coexist with the observer mode, which we will discuss in the following section.

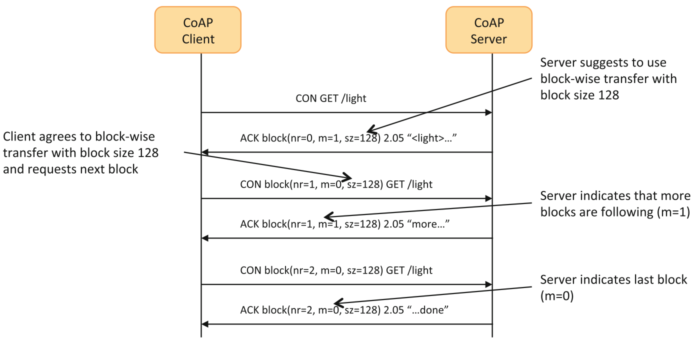


Fig. 9 A CoAP client requests the resource light and receives a block-wise answer. The client requests more blocks until the server informs the client that no more blocks are available

5.1.2 The Observer Mode

Even though CoAP allows retrieving resources efficiently from constrained devices, it is often desirable to monitor a sensor's behavior. For example, given a temperature sensor, it is less efficient to poll the sensor for changes instead of letting the sensor inform the client of a change. Therefore, CoAP offers an extension to synchronize the state of a sensor to a client. This mode is referred to as the observer mode as it follows the idea of the prominent observer pattern as introduced in [22].

Thus, the observer pattern acts like a publish/subscribe system. However, for CoAP, only a best-effort mechanism is specified in [26]. This also means that observing clients are not guaranteed to be in synchronization with the server's state. Nevertheless, Hartke [26] argues that eventually synchronization will be reached.

Clients can register to a resource by including an *observe option* in their request. The observe option itself contains a field that specifies whether this client should be registered or a previous registration should be removed. In case the server supports the observer mode, in addition to serving the request, it includes the observe option as well. For responses, the option indicates that, on the one hand, a registration was successful and, on the other hand, that this message is a notification to the client. The notification includes a number that can be used by the client to determine the order of notifications, e.g., in case an older notification arrives after a newer one. Whenever the resource that the observer subscribed to is changed, the server will issue a notification to the subscribers. To not pose complex state requirements to a server, all notifications may be delivered without acknowledgements (NON-confirmable) allowing the server to dismiss the data from its memory after sending the packet.

Assuming the example of a temperature sensor, it may of course not be of interest to simply know if the temperature increased by a single degree. It is often of interest to monitor if sensors enter or leave a critical region. Such a behavior can be easily implemented with the help of different resources. For example, a server might implement the resource: `coap://server/temperature` that changes whenever the sensor reads another temperature. However, an application programmer might also implement the resource `coap://server/temperature/critical?above=x`, to which an observer may subscribe specifying (using *x*) above which threshold a notification should be sent. The implementation of this logic is up to the application developer and is not inherently a feature of CoAP. Figure 10 illustrates the typical observer pattern using the temperature sensor as an example.

5.1.3 CoAP and Proxies

Especially when looking at the interconnection of factories and cloud services, which are connected to each other over the Internet, an end-to-end UDP channel might be too unreliable to deliver the data as it has to compete with other Internet traffic. Moreover, cloud services may completely lack CoAP support, or CoAP applications may need to interact with existing applications. In addition, the

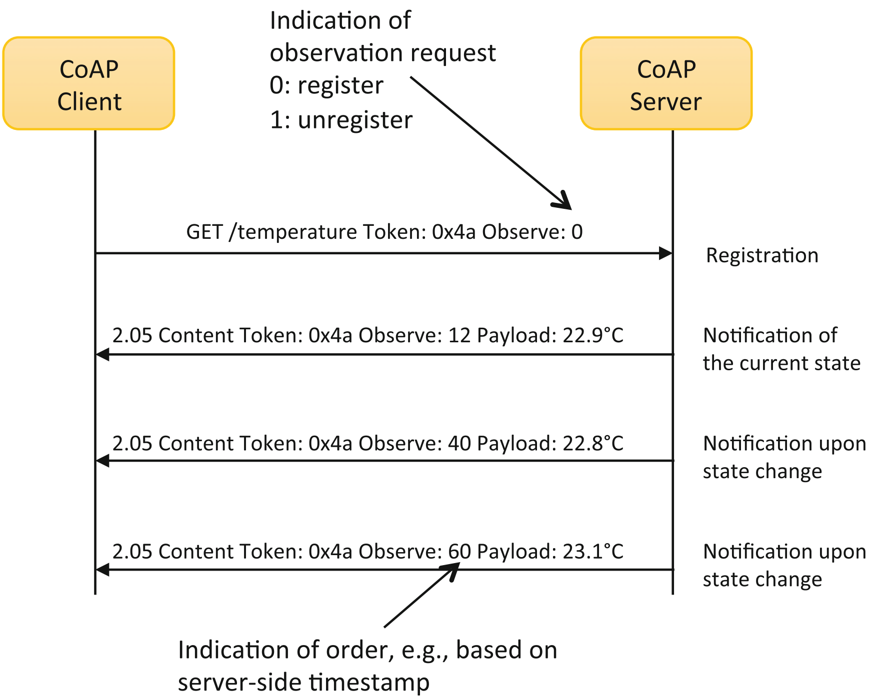


Fig. 10 A CoAP client registers to the resource temperature at a CoAP server. The server replies with several notifications whenever the temperature changes (cf. [26])

constrained nature of the devices employing CoAP may already demand an intermediary that, e.g., caches responses, thereby reducing costly traffic within the factory's network or allowing to access resources of sleeping devices.

Therefore, it might be required to translate between different protocols to reach the CoAP services. As CoAP's semantics are close to HTTP, which is one of the most widely used protocols on the Internet and uses a reliable TCP transport, a proxy at the edge of the factory that translates between HTTP and CoAP is compelling. CoAP's design makes it especially easy to operate an HTTP proxy: CoAP URI's are easily mappable to HTTP URI's by replacing the prefix `coap`: // with `http`: //, still, the path component of the URI needs to map semantically to a CoAP resource. The proxy then issues the CoAP command that is specified in the HTTP request. In addition, the proxy can be used as a cache for popular resources: instead of always fetching the same resource over and over again, the proxy can use the observer pattern to subscribe to the resource and thus relief the factory's network. Of course, what is popular is up to the implementor of the proxy, e.g., resources may be cached after observing that they are requested frequently. In addition, as CoAP is designed for multicast support via IP multicast, a proxy may support mapping a single HTTP request to a multicast request, e.g., if the request is destined to a hostname that maps to a multicast group. In the following, we will highlight

Cloud technology to show how to manage the high number of devices and the huge amount of data that is generated.

5.2 Cloud and Distributed Processing

In the vision of connecting and monitoring all steps and entities in a production process, large amounts of sensor data and many actuators are exposed. To manage these, cloud solutions to store, process and act upon the data have been proposed [10, 29, 40]. Cloud solutions offer a scalable and cost-efficient way of managing the data while maintaining reliability [3]. Thus, it is not required to operate a data center to store and process all data: Cloud solutions allow seamless scaling when the demands increase. However, it must be considered that in these Platform as a Service (PaaS) or Software as a Service (SaaS) deployments, the data may potentially be stored and processed on machines operated by others. So, the privacy and security of the data should be considered at least when the data leaves its home network. To this end, many platforms have been proposed that especially tackle the challenges in IoT-like deployments. For example, Xively [41] focuses on a connected product management, offering libraries and SDKs for a rapid development and interconnection of things. Numerous other platforms exist, each offering different capabilities, different payment models or even open source solutions: For example, Al-Fuqaha et al. [3] offer a short comparison of different cloud platforms. Many cloud platforms not only contain ready-made libraries for interconnecting things and the cloud, in addition, they contain tools and engines to process and visualize the gathered data. Still, a need for common standards that allow to better specify the integration with the cloud is desirable. This would enable a much simpler move between different vendors and platforms.

6 Conclusion and Outlook

In this chapter, we discuss the challenges of an interconnected IIoT and give possible solutions tackling these challenges throughout the protocol stack. The demand for wireless solutions poses a tough problem that challenges the delay and guarantees that can be given on the physical and data link layer. The pursuit of cheap but flexible hardware poses an enormous challenge to network and communication. The constrained nature of these devices forbids the use of well-established and heavily proven Internet technologies. Thus, several standards such as WirelessHART, ISA110 or 6LoWPAN have emerged that tackle these problems. Which technology to use heavily depends on the usage scenario, the demands, and the deployed devices. Nevertheless, standardization is a key enabling factor for interoperability and allows developing and deploying future-proof systems. The same demands are posed to application layer protocols: The constrained nature might make it impossible or infeasible to work with established protocols. Protocols such as CoAP or MQTT offer application programmers well-known paradigms such as publish-subscribe and framing the application data in a meaningful fashion. Still, while these protocols work well in monitoring and soft control environments, their use in delay-sensitive environments remains a challenge. By interfacing local sensors and actuators with cloud computation and aggregation, new monitoring capabilities and control algorithms become feasible. In the end, the choice for a cloud platform heavily depends on the application scenario and should be evaluated in terms of supporting technologies and scalability, as well as its capabilities to process and visualize gathered data.

References

1. Akerberg J, Gidlund M, Bjorkman M (2011) Future research challenges in wireless sensor and actuator networks targeting industrial automation. In: 9th IEEE international conference on industrial informatics (iNDIN), pp 410–415. doi:10.1109/INDIN.2011.6034912
2. Akyildiz IF, Wang X (2005) A survey on wireless mesh networks. IEEE Commun Mag 43 (9):23–30. doi:10.1109/MCOM.2005.1509968
3. Al-Fuqaha A, Guizani M, Mohammadi M, Aledhari M, Ayyash M (2015) Internet of Things: a survey on enabling technologies, protocols and applications. IEEE Commun Surv Tutor 17 (4):2347–2376. doi:10.1109/COMST.2015.2444095
4. Andreev S, Galinina O, Pyattaev A, Gerasimenko M, Tirronen T, Torsner J, Sachs J, Dohler M, Koucheryavy Y (2015) Understanding the IoT connectivity landscape: a contemporary M2M radio technology roadmap. IEEE Commun Mag 53(9):32–40. doi:10.1109/MCOM.2015.7263370
5. Atzori L, Iera A, Morabito G (2010) The Internet of Things: a survey. Comput Netw 54 (15):2787–2805. doi:10.1016/j.comnet.2010.05.010
6. Basagni S, Conti M, Giordano S, Stojmenovic I (2004) Mobile ad hoc networking. Wiley, Piscataway
7. Bormann C (2014) 6LoWPAN-GHC: generic header compression for IPv6 over low-power wireless personal area networks (6LoWPANs). RFC 7400; RFC Editor
8. Bormann C, Shelby Z (2015) Block-wise transfers in CoAP. IETF Secretariat; draft-ietf-core-block-18.txt, Working Draft
9. Bormann C, Castellani A, Shelby Z (2012) CoAP: an application protocol for billions of tiny internet nodes. IEEE Internet Comput 16(2):62–67. doi:10.1109/MIC.2012.29
10. Botta A, de Donato W, Persico V, Pescape A (2014) On the integration of cloud computing and Internet of Things. In: International conference on future Internet of Things and cloud (FiCloud), pp 23–30. doi:10.1109/FiCloud.2014.14
11. Clausen T, Jacquet P (2003) Optimized link state routing protocol (OLSR). RFC 3626; RFC Editor
12. Colitti W, Steenhaut K, De Caro N (2011) Integrating wireless sensor networks with the web. In: Proceedings of workshop on extending the internet to low power and lossy networks (IP+SN)
13. Cutler TR (2014) The internet of manufacturing things. Ind Eng 46(8):37–41
14. Deering S, Hinden R (1998) Internet protocol, version 6 (IPv6) specification. RFC 2460; RFC Editor
15. Diggavi S, Al-Dhahir N, Stamoulis A, Calderbank A (2004) Great expectations: the value of spatial diversity in wireless networks. Proc IEEE 92(2):219–270. doi:10.1109/JPROC.2003.821914

16. Dinh NQ, Kim D-S (2012) Performance evaluation of priority CSMA-CA mechanism on ISA100.11a wireless network. Comput Stand Interfaces 34(1):117–123. doi:10.1016/j.csi.2011.06.001
17. Dombrowski C, Gross J (2015) EchoRing: a low-latency, reliable token-passing MAC protocol for wireless industrial networks. In: Proceedings of 21th European wireless conference, pp 1–8
18. Fielding RT, Taylor RN (2000) Principled design of the modern web architecture. In: Proceedings of the 22nd international conference on software engineering. ACM, pp 407–416. doi:10.1145/337180.337228
19. Fielding R, Gettys J, Mogul J, Frystyk H, Berners-Lee T (1997) Hypertext transfer protocol—HTTP/1.1. RFC 2616; RFC Editor
20. Foschini G, Chizhik D, Gans M, Papadias C, Valenzuela R (2003) Analysis and performance of some basic space-time architectures. IEEE J Sel Areas Commun 21(3):303–320. doi:10.1109/JSAC.2003.809456
21. Frotzscher A, Wetzker U, Bauer M, Rentschler M, Beyer M, Elspass S, Klessig H (2014) Requirements and current solutions of wireless communication in industrial automation. In: IEEE international conference on communications workshops (ICC 2014), pp 67–72. doi:10.1109/ICCW.2014.6881174
22. Gamma E, Helm R, Johnson R, Vlissides J (1995) Design patterns: elements of reusable object-oriented software. Addison-Wesley Longman Publishing Co. Inc., Boston
23. Groover MP (2007) Automation, production systems, and computer-integrated manufacturing, 3rd edn. Prentice Hall Press, Upper Saddle River
24. Gummalla ACV, Limb JO (2000) Wireless medium access control protocols. IEEE Commun Surv Tutor 3(2):2–15. doi:10.1109/COMST.2000.5340799
25. Han S, Zhu X, Mok AK, Chen D, Nixon M (2011) Reliable and real-time communication in industrial wireless mesh networks. In: 17th IEEE real-time and embedded technology and applications symposium (RTAS). IEEE, pp 3–12. doi:10.1109/RTAS.2011.9
26. Hartke K (2014) Observing resources in the constrained application protocol (CoAP). RFC 7641; RFC Editor
27. Huang S, Zhang Z (2011) Principles of FECs with evaluating different types of FEC used in the internet and wireless networks. In: International conference on electronics, communications and control (ICECC), pp 2181–2184. doi:10.1109/ICECC.2011.6066671
28. Hui J, Thubert P (2011) Compression format for IPv6 datagrams over IEEE 802.15.4-based networks. RFC 6282; RFC Editor
29. Hummen R, Henze M, Catrein D, Wehrle K (2012) A cloud design for user-controlled storage and processing of sensor data. In: IEEE 4th international conference on cloud computing technology and science (CloudCom), pp 232–240. doi:10.1109/CloudCom.2012.6427523
30. Hunkeler U, Truong HL, Stanford-Clark A (2008) MQTT-S—a publish/subscribe protocol for wireless sensor networks. In: 3rd international conference on communication systems software and middleware and workshops (COMSWARE 2008), pp 791–798. doi:10.1109/COMSWA.2008.4554519
31. IEC (2010) Industrial communication networks—wireless communication network and communication profiles—WirelessHART. IEC 62591 Ed. 1.0 b:2010 Standard
32. IEEE (2005) IEEE standard for information technology—local and metropolitan area networks—specific requirements—Part 15.1a: wireless medium access control (MAC) and physical layer (PHY) specifications for wireless personal area networks (WPAN). IEEE Std 802151-2005 (Revision of IEEE Std 802151-2002). doi:10.1109/IEEESTD.2005.96290
33. IEEE (2012) IEEE standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements Part 11: wireless lan medium access control (MAC) and physical layer (PHY) specifications. IEEE Std 80211-2012 (Revision of IEEE Std 80211-2007). doi:10.1109/IEEESTD.2012.6178212
34. IEEE (2012) IEEE standard for ethernet. IEEE Std 8023-2012 (Revision to IEEE Std 8023-2008). doi:10.1109/IEEESTD.2012.6419735

35. IEEE (2006) IEEE standard for local and metropolitan area networks—Part 15.4: low rate wireless personal area networks (LR-WPANs). IEEE Std 802154-2006 (Revision of IEEE Std 802154-2003). doi:10.1109/IEEESTD.2006.232110
36. ISA (2011) Wireless systems for industrial automation: process control and related applications
37. Ishaq I, Carels D, Teklemariam GK, Hoebeke J, den Abeele FV, Poorter ED, Moerman I, Demeester P (2013) IETF standardization in the field of the Internet of Things (IoT): a survey. J Sens Actuator Netw 2(2):235–287
38. Kushalnagar N, Montenegro G, Schumacher C (2007) IPv6 over low-power wireless personal area networks (6LoWPANs): overview, assumptions, problem statement, and goals. RFC 4919; RFC Editor
39. Laneman J, Tse D, Wornell GW (2004) Cooperative diversity in wireless networks: efficient protocols and outage behavior. IEEE Trans Inf Theor 50(12):3062–3080. doi:10.1109/TIT.2004.838089
40. Li F, Voegler M, Claessens M, Dustdar S (2013) Efficient and scalable IoT service delivery on cloud. In: IEEE sixth international conference on cloud computing (CLOUD), pp 740–747. doi:10.1109/CLOUD.2013.64
41. LogMeIn Inc. (2015) Xively by LogMeIn. https://xively.com. Accessed 20 May 2016
42. Lu G, Krishnamachari B, Raghavendra C (2004) Performance evaluation of the IEEE 802.15.4 MAC for low-rate low-power wireless networks. In: IEEE international conference on performance, computing, and communications, pp 701–706. doi:10.1109/PCCC.2004.1395158
43. Marzetta T (2010) Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans Wirel Commun 9(11):3590–3600. doi:10.1109/TWC.2010.092810.091092
44. Montenegro G, Kushalnagar N, Hui J, Culler D (2007) Transmission of IPv6 packets over IEEE 802.15.4 networks. RFC 4944; RFC Editor
45. Mulligan G (2007) The 6LoWPAN architecture. In: Proceedings of the 4th workshop on embedded networked sensors (EmNets 2007). ACM, pp 78–82, doi:10.1145/1278972.1278992
46. Neumann P (2007) Communication in industrial automation—what is going on? Control Eng Pract 15(11):1332–1347. doi:10.1016/j.conengprac.2006.10.004
47. Nixon M (2012) A comparison of WirelessHART™ and ISA100. 11a
48. Nobre M, Silva I, Guedes LA (2015) Routing and scheduling algorithms for WirelessHARTNetworks: a survey. Sensors 15(5):9703–9740. doi:10.3390/s150509703
49. Perkins C, Belding-Royer E, Das S (2003) Ad hoc on-demand distance vector (AODV) routing. RFC 3561; RFC Editor
50. Petersen S, Carlsen S (2011) WirelessHART versus ISA100.11a: the format war hits the factory floor. IEEE Ind Electron Mag 5(4):23–34. doi:10.1109/MIE.2011.943023
51. Phoenix Contact (2000) INTERBUS RS232 communication manual
52. Pister K, Thubert P, Dwars S, Phinney T (2009) Industrial routing requirements in low-power and lossy networks. RFC 5673; RFC Editor
53. Poor HV, Wornell GW (1998) Wireless communications: signal processing perspectives. Prentice Hall PTR, Upper Saddle River
54. Popovski P (2014) Ultra-reliable communication in 5G wireless systems. In: 1st international conference on 5G for ubiquitous connectivity (5GU), pp 146–151. doi:10.4108/icst.5gu.2014.258154
55. Postel J (1980) User datagram protocol. RFC 768; RFC Editor
56. Postel J (1981) Transmission control protocol. RFC 793; RFC Editor
57. Qiao D, Choi S, Shin KG (2007) Interference analysis and transmit power control in IEEE 802.11a/h wireless LANs. IEEE/ACM Trans Netw 15(5):1007–1020. doi:10.1109/TNET.2007.900381
58. Serror M, Dombrowski C, Wehrle K, Gross J (2015) Channel coding versus cooperative ARQ: reducing outage probability in ultra-low latency wireless communications. In: IEEE GlobeCom workshop on ultra-low latency and ultra-high reliability in wireless communications

59. Sha M, Gunatilaka D, Wu C, Lu C (2015) Implementation and experimentation of industrial wireless sensor-actuator network protocols. In: Wireless sensor networks. Springer, Switzerland, pp 234–241. doi:10.1007/978-3-319-15582-1_15
60. Shelby Z, Bormann C (2011) 6LoWPAN: the wireless embedded internet. Wiley, Chichester
61. Shelby Z, Hartke K, Bormann C (2014) The constrained application protocol (CoAP). RFC 7252; RFC Editor
62. Shelby Z, Chakrabarti S, Nordmark E, Bormann C (2012) Neighbor discovery optimization for IPv6 over low-power wireless personal area networks (6LoWPANs). RFC 6775; RFC Editor
63. Song J, Han S, Mok A, Chen D, Lucas M, Nixon M (2008) WirelessHART: applying wireless technology in real-time industrial process control. In: IEEE real-time and embedded technology and applications symposium (RTAS 2008), pp 377–386. doi:10.1109/RTAS.2008.15
64. Spencer Q, Peel C, Swindlehurst A, Haardt M (2004) An introduction to the multi-user MIMO downlink. IEEE Commun Mag 42(10):60–67. doi:10.1109/MCOM.2004.1341262
65. Stanford-Clark A, Truong HL (2013) MQTT for sensor networks (MQTT-SN) protocol specification version 1.2. http://mqtt.org/documentation. Accessed 20 May 2016
66. Suriyachai P, Roedig U, Scott A (2012) A survey of MAC protocols for mission-critical applications in wireless sensor networks. IEEE Commun Surv Tutor 14(2):240–264. doi:10.1109/SURV.2011.020211.00036
67. Tarokh V, Seshadri N, Calderbank A (1998) Space-time codes for high data rate wireless communication: performance criterion and code construction. IEEE Trans Inf Theor 44(2):744–765. doi:10.1109/18.661517
68. Thomesse J-P (2005) Fieldbus technology in industrial automation. Proc IEEE 93(6):1073–1101. doi:10.1109/JPROC.2005.849724
69. Thomson S, Narten T, Jinmei T (2007) IPv6 stateless address autoconfiguration. RFC 4862; RFC Editor
70. Tovar E, Vasques F (1999) Real-time Fieldbus communications using profibus networks. IEEE Trans Ind Electron 46(6):1241–1251. doi:10.1109/41.808018
71. Vasseur J-P, Dunkels A (2010) Interconnecting smart objects with IP: the next internet. Morgan Kaufmann, Burlington
72. Vincent SJ (2001) Foundation Fieldbus high speed ethernet control system. Fieldbus Inc
73. Watteyne T, Mehta A, Pister K (2009) Reliability through frequency diversity: why channel hopping makes sense. In: Proceedings of the 6th ACM symposium on performance evaluation of wireless ad hoc, sensor, and ubiquitous networks. ACM, pp 116–123. doi:10.1145/1641876.1641898
74. Willig A, Kubisch M, Hoene C, Wolisz A (2002) Measurements of a wireless link in an industrial environment using an IEEE 802.11-compliant physical layer. IEEE Trans Ind Electron 49(6):1265–1282. doi:10.1109/TIE.2002.804974
75. Willig A, Matheus K, Wolisz A (2005) Wireless technology in industrial networks. Proc IEEE 93(6):1130–1151. doi:10.1109/JPROC.2005.849717
76. Winter T, Thubert P, Brandt A, Hui J, Kelsey R, Levis P, Pister K, Struik R, Vasseur J, Alexander R (2012) RPL: IPv6 routing protocol for low-power and lossy networks. RFC 6550; RFC Editor
77. Wirtz H, Rüth J, Serror M, Zimmermann T, Wehrle K (2015) Enabling ubiquitous interaction with smart things. In: 12th annual IEEE international conference on sensing, communication, and networking (IEEE SECON 2015), pp 256–264
78. Wu D, Greer MJ, Rosen DW, Schaefer D (2013) Cloud manufacturing: strategic vision and state-of-the-art. J Manuf Syst 32(4):564–579. doi:10.1016/j.jmsy.2013.04.008
79. Yazar D, Dunkels A (2009) Efficient application integration in IP-based sensor networks. In: Proceedings of the first ACM workshop on embedded sensing systems for energy-efficiency in buildings. ACM, pp 43–48. doi:10.1145/1810279.1810289

Communications for Cyber-Physical Systems

Mohammad Elattar, Verena Wendt and Jürgen Jasperneite

1 Introduction

A CPS, as shown in Fig. 1, consists of one or more interconnected autonomous components or units where services of each unit are visible to the other units of the system. One of the distinctive features of such systems is networking, which provides information exchange not only at local levels within each unit of the CPS but also at higher levels between the units. The concept of combining computing and physical processes has been already considered in engineered systems. Such systems have existed since a few decades and are usually called "embedded systems". Examples of embedded systems include home appliances, aircraft control systems, and automotive electronics. The main difference between embedded systems and CPSs is that embedded systems mostly represent black boxes. They do not show their computing capability to the outside and no outside connectivity can alter their software behavior [37]. In a CPS, the units of the system are feature-rich, networked, and cooperate together using communication networks. As a result, communications infrastructure of a CPS can be compared to the nervous system in humans that connects and coordinates the different body parts. Local control networks are usually

M. Elattar (✉) · V. Wendt · J. Jasperneite
inIT – Institute Industrial IT, OWL University of Applied Sciences,
Langenbruch 6, 32657 Lemgo, Germany
e-mail: mohammad.elattar@stud.hs-owl.de

V. Wendt
e-mail: verena.wendt@hs-owl.de

J. Jasperneite
e-mail: juergen.jasperneite@hs-owl.de; juergen.jasperneite@iosb-ina.fraunhofer.de

J. Jasperneite
Fraunhofer IOSB-INA Application Center Industrial Automation,
Langenbruch 6, 32657 Lemgo, Germany

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series
in Wireless Technology, DOI 10.1007/978-3-319-42559-7_13

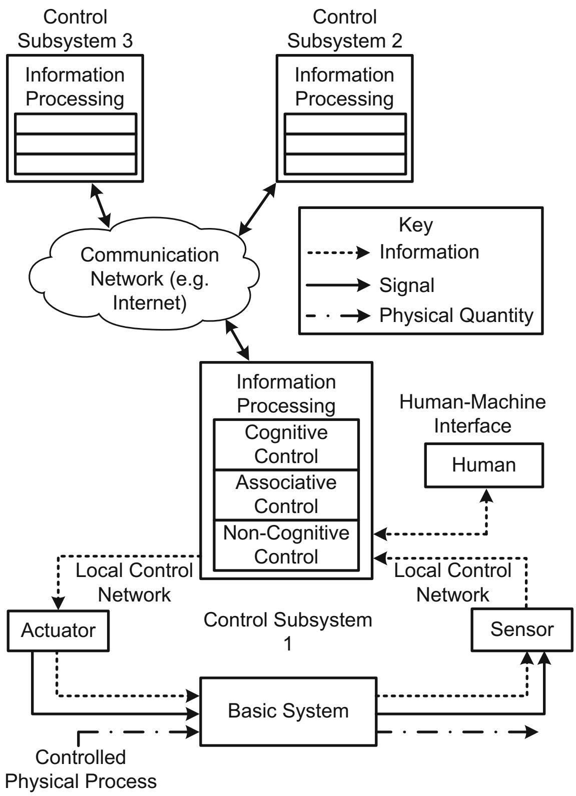


Fig. 1 General architecture of CPSs [53]

used to realize communications within each unit of a CPS to achieve real-time communications required by the technical processes under control. In contrast, the real-time requirements between the units of the CPS are usually less stringent and, consequently, other types of networks including the Internet might be used to realize the communications.

As illustrated in Fig. 1, a CPS unit generally consists of the following entities: a computational entity to monitor and control the physical process, sensing and actuating entities to interact with the physical process, and the physical process to be controlled. The use of a communication network to connect these components within each unit and also the units with each other brings several advantages including remote monitoring and control, scalability, flexibility, and decentralization of control [23]. In industrial applications, for example, a variable number of distributed components including sensors, actuators, and controllers, which might be located far away from each other and belong to the same or different applications, share a single communication network to exchange control information.

2 Data Communication Networks

In general, a communication network is a system that allows two or more endpoints (also called hosts) to be connected and to exchange data. The term endpoint does not necessarily refer to computers only, rather to any kind of equipment that is able to connect to the network. The network itself, as shown in Fig. 2, consists of nodes and links to connect them. The endpoints connect to the network by connecting to some of the network nodes. If we assumed that all links in the network are bidirectional, then, each node is capable of receiving or forwarding data over any of the connected links. A node can receive data from an endpoint or from another node. Similarly, a node can forward data to an endpoint or another node. The physical medium used to realize the links might be wired (e.g. copper wire or optical fiber) or wireless (e.g. microwave radio transmission) and might differ from one link to another. In addition, each of the links might have different capacity that is measured by the maximum bit rate provided.

In this chapter, we consider only data communication networks where the information to be carried consists of 0 and 1 data streams. Analog communication networks, such as analog telephone networks, where analog signals are transferred without digital encoding are not considered. This is basically due to the digital nature of CPSs.

In data networks, data are carried over the network in small units, called packets, which have certain formats determined by the network. The network also specifies the maximum size of the packets and the extra information (beside the actual data) needed to transfer them over it. The extra information includes the source and destination endpoints addresses and the number of bytes. Packets usually consists of a header where the extra information is included and a payload where the actual information is included. Based on the network, a packet might also include a trailer to carry part of the extra information.

The widely adopted classification of data networks is based on the area covered and the number of users served by the network. According to this classification, there are local-area networks (LANs) and wide-area networks (WANs). LANs refer to networks that are confined in space such as those in a single building or in a campus. In contrast, WANs refer to networks that span large geographical areas and

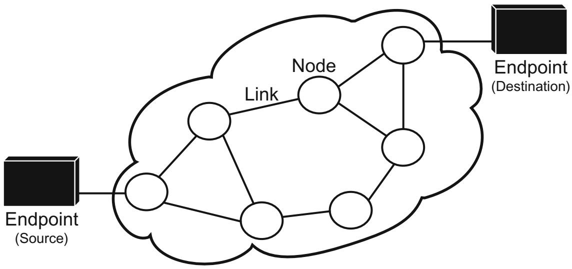


Fig. 2 General topology of a communications network

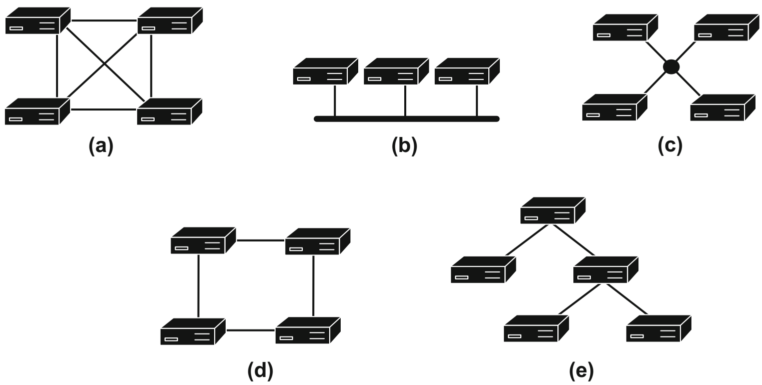


Fig. 3 Fundamental topologies of communications network: **a** full-mesh, **b** bus, **c** star, **d** ring, and **e** tree

connect two or more LANs. An example of a WAN is the Internet which is considered as the largest WAN.

Another method to classify data networks is based on the physical topology of their components. Among the possible organizations of the physical topology are full-mesh, star, bus, ring, and tree. These topologies are illustrated in Fig. 3. In a full-mesh topology, every host connects to every other host in the network. This topology provides high redundancy and performance, but it can be used only when the number of hosts is small. In a bus topology, a shared medium such as a cable is used to connect all hosts. Data sent by any host on the bus are received by all other hosts. Physical damages to the bus usually divide the network and isolate its different parts. As a result, bus topology is usually difficult to maintain. A widely adopted physical organization of networks is the star topology where each host in the network uses a separate link to connect to a central entity. The topology provides higher flexibility with regard to adding or removing hosts, however, a failure of the central entity results in a failure of the entire network. In a ring topology, hosts are connected in a circular fashion where each host has two neighbors. In this organization, the data travel in one direction (clockwise or counter clockwise) around the ring. Each host on the ring acts as a repeater and forwards the data to the next hop till it reaches the designated destination. A failure of a host or a link between two hosts will result in a failure of the network. Lastly, the tree topology divides the network into levels. The hosts at the lowest level of the tree, known as the leaves, can act as senders or receivers while hosts at higher levels also act as repeaters. This topology is usually used to provide cost-effective organization to connect large number of hosts (leaves). The above mentioned topologies are fundamental topologies where real networks usually combine them. For example, an Internet service provider (ISP) might adopt the ring topology for the core part of the network and the tree topology for last mile connectivity.

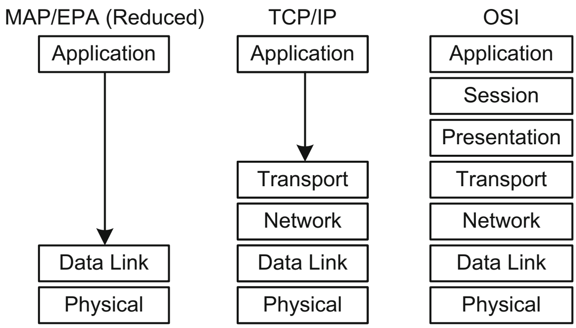


Fig. 4 Protocol stacks of OSI reference model, TCP/IP, and reduced MAP/EPA model

All kinds of communication including human face-to-face communication and network communication need predetermined rules in order to be successful. In data networks, protocols organize the different tasks between two communicating devices. For example, they define the format and maximum size of data packets, the way to begin and end communication between two hosts, and the way packets are routed between hosts through the data network. At the beginning of networking industry, manufacturers provided proprietary equipment and protocols for networking. As the cooperation between the different companies started to increase, the need for sharing data and, consequently, networks increased, too. As a result, standards for networking became necessary to provide interoperability between vendors [13]. In particular, the Open System Interconnection (OSI) reference model and the Transmission Control Protocol/Internet Protocol (TCP/IP) model were created. The protocol stacks of the OSI and TCP/IP models are illustrated in Fig. 4. As we can see in the figure, both models are layered models to tackle the complexity of describing network communications. This is also the adopted approach in implementing the communication process in real networks. Each layer in these models provides one or more services to the layer above it. Therefore, the different protocols that perform specific functions or tasks in the entire communication process are grouped into the different layers of these reference models. In contrast to the OSI model which describes the communication process in general, the TCP/IP model considers only the TCP/IP protocol suite and its communication process. A brief description for each layer of the OSI model is given as follows:

Application layer: Provides services for user applications to access the network.

Presentation layer: Provides information to the application layer with regard to the data format.

Session layer: Establishes, manages, and terminates sessions between user applications.

Transport layer: Performs data segmentation and numbering at the source, data transfer, and data reassembly at the destination.

Network layer: Creates data packets and addresses them for end-to-end delivery in a multi-node network.

Data link layer: Creates data frames and address them for delivery between nodes that share a physical layer.

Physical layer: Transmits and receives binary data symbols over a physical media.

3 Types of Communication Networks for CPSs

LANs and WANs can be categorized, based on the domain, to industrial and general purpose networks. The majority of communications in today's CPSs are realized using industrial communication networks to fulfill the real-time communication requirements. Local industrial control networks are used extensively within the control subsystems of a CPS. The motivation for such networks is to replace the point-to-point communication at the plant level between the different field devices (e.g. sensors and actuators) and their corresponding controllers (e.g. programmable logic controllers (PLCs)) by multipoint-to-multipoint communication (bus). Local industrial networks are usually confined in space and differ from general purpose networks as the requirements of such networks are much higher. As a result, they are capable of providing real-time communication, predictable throughput, and very low down times, and can operate in harsh environments (e.g. high noise environments). In addition, the data packets over such networks are characterized by having small sizes with low protocol overhead and the topology of such networks can take different forms (e.g. star, ring, tree, etc.) depending on the application. Compared to the OSI network model with 7 layers, most of the local industrial networks are based on the Manufacturing Automation Protocol/Enhanced Performance Architecture (MAP/EPA) reduced network model. As shown in Fig. 4, the MAP/EPA model consists only of the application, data-link and physical layers [18].

A wide variety of local industrial networks were developed over the last few decades, called fieldbus systems. These systems, with the majority standardized in the IEC 61158 standard, were proposed for different industrial markets and offer different features. Examples include Controller Area Network (CAN), PROcess-FIeld Bus (PROFIBUS), INTERBUS, and Factory Instrumentation Protocol (FIP). In recent years, newer Ethernet-based fieldbus standards were proposed such as Ethernet for Control Automation Technology (EtherCAT), Process Field Net (PROFINET), or Ethernet/IP. This is mainly due to the technological advances in Ethernet which allowed new features including real-time communication capabilities, high data rates, full-duplex data transmission, and low congestion with the use of switched networks.

Existing industrial WAN networks today are mainly used by the Supervisory Control and Data Acquisition (SCADA) systems to monitor and control remote industrial infrastructures. Such systems usually consist of remote field devices such as remote terminal units (RTU) or PLCs that connect the remote components (e.g. sensors) to the WAN network of the enterprise (e.g. utility operator). The field devices at the remote sites connect in their turn to a central supervisory computer that connects to a human-machine interface. The intermediate layer of field devices between sensors and communication infrastructure allows the digital transmission

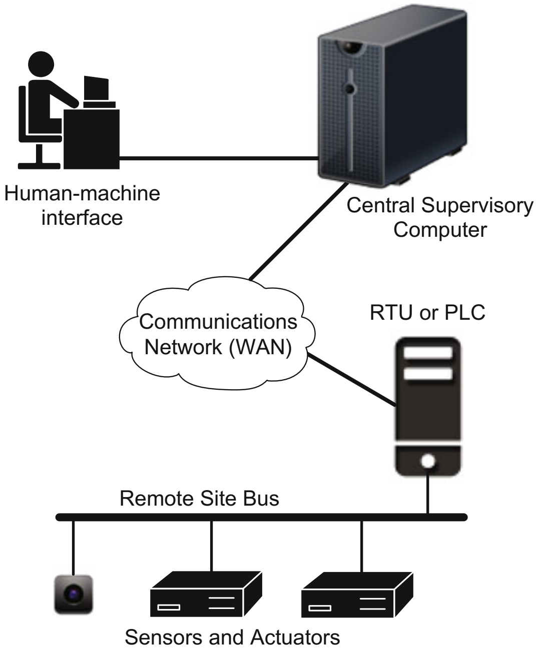


Fig. 5 Simplified diagram for remote site monitoring using SCADA

of sensor signals using industrial communication protocols. Almost all SCADA WANs are using private networks (fiber-optic or radio links built specially between the sites of the system) or dedicated leased lines from national WAN operators to connect the different remote sites of the enterprise. Figure 5 shows a simplified view of SCADA system for remote site monitoring. For simplicity, additional components such as modems were not included in the figure.

One example of the common SCADA protocols used between the field devices and the master controller represented by the central supervisory computer is the Distributed Network Protocol (DNP3) [11] standardized in the IEEE 1815 standard [26]. DNP3 allows the master controller to poll field devices for measurement and status information either periodically or on demand and also to send commands to these devices. Similar to industrial LANs protocols, only the application, transport, and data link layers of the OSI model are present in DNP3 with a maximum transport protocol data unit (PDU) of 250 bytes. The protocol is defined on serial connections at the physical layer (e.g. using RS232 standard) and supports IP-based networks by adding a fourth layer below the data link layer. This layer is called the Data Connection Management and allows the DNP3 with its layers to be the application layer of popular transport protocols such as the TCP Protocol and the User Datagram Protocol (UDP).

Compared to local industrial networks, SCADA systems are usually considered as slow response systems where the response might take several seconds and/or requires the intervention of a human operator [9]. The response time of SCADA systems is increased due to several facts: The nature of existing SCADA protocols such as DNP3, the need for special hardware to allow their transmission over prominent WAN networks (e.g. IP-based networks), and the use of limited bandwidth links to reduce networking expenses. As a result, SCADA systems are usually considered to perform coordination rather than direct real-time process control.

Due to the unpredictable timing behavior of general purpose LANs, they are not used at the field level of control between PLCs, sensors, and actuators. However, such networks might be used at the supervisory level to monitor and coordinate the different processes within a plant. In this level, the real-time and determinism requirements are much less stringent. Similarly, general purpose WANs are also used for CPS applications with low real-time and reliability requirements, but with the components distributed over the large geographical area [9]. Moreover, new wireless standards such as the Long-Term Evolution (LTE) provide very low latencies (as low as 50 ms), high data rates, and very comprehensive framework for QoS. This increased the interest in utilizing them for CPSs instead of using private WANs. However, the existing usages of such technology have several drawbacks that limit their benefits to CPSs. Such drawbacks will be detailed later in this chapter. It is also worth mentioning here that upcoming wireless communication standards, denoted by the 5th generation (5G) [43], consider the communication requirements of CPSs. More particularly, the requirements of 5G networks as defined by the Next Generation Mobile Networks (NGMN) Alliance include support for enormous number of concurrent connections (e.g. to support large deployments of sensors), very low end-to-end latency in the order of 1 ms, enhanced coverage, and improved spectral and signaling efficiencies compared to the 4th generation technologies (e.g. LTE).

CPSs were envisioned to use the Internet as their future communication network. This is also motivated by the technological advances in communication technologies that enabled high-speed data communications. However, Internet utilization for CPSs is limited by factors such as the high heterogeneity of commercial networks constituting it, the different operator and QoS policies, and the insufficient communication reliability provided by it [50, 56]. More details with this regards are provided in a later section of this chapter. A future trend that is also expected to enable the realization of Internet-based CPSs is the Tactile Internet [27] that is characterized by having very low latency and high reliability, availability, and security. With such characteristics, the Tactile Internet will support not only existing applications but also new ones in a wide range of domains including industrial automation, healthcare, transportation systems, and gaming. Nevertheless, Tactile Internet imposes very stringent requirements on the communication infrastructure that might not be supported by existing technologies and demand the development of capable ones.

4 Impact of Communication Network Deficiencies

Data networks exhibit by necessity performance limitations and reliability limitations. Performance limitations are mainly caused by the nature of the physical media used. By contrast, reliability limitations are attributed to several reasons including components and network failures (caused by, for example, procedural errors and software or hardware updates), oversubscription, environmental conditions (e.g. effect of weather on wireless communications), slow recovery of routing protocols when communication paths fail, and power disruptions. As a result, some deficiencies with regard to communications performance and reliability arise. The major performance deficiencies are described first in the following:

Time delay: Which is the average time required for the delivery of data packets between the source and destination endpoints in the network. This time delay depends on several factors including (1) the media access scheme which determines the time taken to accept a new packet by the network; (2) the transmission time of packets inside the network. The later factor depends in turn on the propagation time of signals over the network medium and the queuing and processing times of network components.

Jitter: If the time delay introduced by the communication network is variable, the arrival time of packets will fluctuate from one packet to another, which is known as jitter. This is mainly attributed to the highly varying nature of network conditions such as network load (e.g. congestion) and the quality of communications channels or links.

Packet loss: Another communications deficiency is the loss of data packets due to several reasons including transmission errors in error prone channels (e.g. wireless channels) and buffers overflow of network devices during congestion.

Limited bandwidth: Communications channels of data networks have finite capacity, which is mainly caused by the limited capacity of physical media used. As a result, the data transfer rate over communication networks is also limited. This imposes a constraint on overall control system operation, especially, when considering that such communication networks are shared between different components and applications.

The above-mentioned performance deficiencies were considered in [35] to be sufficient metrics to describe the provided quality of service (QoS) by a computer network (network performance). Where QoS, as defined in ETR003 ETSI [15, 28], is the user satisfaction determined by the collective impact of service performance. Such QoS is practically represented by a set of performance metrics such as average time delay and provided data rate to give a mean to specify required performance. It was also indicated in [31] that improving QoS, in general, requires minimizing the time delay. While the effect of the other metrics such as jitter can be reduced utilizing existing approaches that provide a tradeoff between these metrics and time delay. Therefore, and due to the limited scope, we consider only the impacts of time delay in more details.

As CPSs incorporate different control systems, the time delay in the control loops of these systems has a significant impact on control performance. Here, system stability is a key control performance parameter and depends on the response time of the system defined as the maximum time allowed between the occurrence of an event and applying the corresponding reaction [58]. When a control loop is closed using a communication network, then this will entail performance degradation or even destabilization of the system [54] due to the presence of time delay and other communications deficiencies. In this context, the traffic of CPSs, even between their units, is usually characterized as a real-time (RT) traffic [21]. The notion of RT traffic means that the traffic has some sort of an upper bound or deadline on information delivery delay. This notion can be further featured, depending on the effect of the deadline violation on the control system, to be either soft or hard. If missing the deadline will mark the late information as useless or even negatively impact correct system operation, then the deadline is hard. If otherwise the late information will degrade performance efficiency of the system without jeopardizing its operation correctness, then the deadline is soft. Consequently, RT systems are usually classified to be soft RT or hard RT systems.

The importance of time delay for systems performance can also be observed from the required network performance for CPSs. Over the last decade, many CPSs applications have been proposed along with studies estimating their traffic characteristics and/or communication requirements [7, 33, 56]. In these studies, time delay was considered as the key performance metric of communication networks in order to realize the proposed CPSs. Moreover, it was clear that all proposed CPS applications require an upper bound on the communication delay rather than a fixed value. Consequently, the effect of jitter on such systems was not considered as long as the information delivery deadline is not violated. Indications on the needed data rates for such CPS applications were also provided. Other network performance metrics such as maximum allowed packet error and loss rates were not considered in many of these studies. However, the presence of such communications deficiencies will certainly degrade CPSs performance [23].

Similarly, reliability deficiencies of communication networks can also negatively impact control loops stability or even stop their operations. Network reliability is defined as the probability that the network and its components will be satisfactorily operational for a specific time interval [40]. This definition also follows the general definition of system reliability [46]. The attributes of system reliability include availability and maintainability [5]. System availability refers to the probability that the system is at a failure-free state at the instant of time when it is first accessed. On the other hand, maintainability refers to the capability of the system to adjust and correct faults, ameliorate performance or adapt to a varied environment. The terms reliability and availability are usually used interchangeably [9, 40]. Therefore, it is useful to indicate their relationship in brief. The reliability definition implies successful operation over a period of time whereas availability definition implies successful operation at the instant of time when the system service is requested. In the literature, several works also considered the relationship between reliability and availability. For example, McCabe [42] indicated that unavailable system is originally not fulfilling the specified requirements to be considered reliable. In addition, both McCabe [42] and Al-Kuwaiti et al. [5] used availability as a measure for system reliability. This also coincides with how the desired communication networks reliability for many CPSs is specified [9, 56]. As an example, 99.99 % reliability means that the probability of network unavailability is 0.0001. Such unavailability probability is informally expressed as the duration of time in a year that the network will be unavailable. More specifically, 99.99% reliability means that the unavailability of the network will be less than one hour (0.0001∗365∗24∗60 min) in a one-year interval.

Another important issue regarding communication reliability is the data transfer reliability which refers to the delivery of messages to the intended recipient(s) complete, uncorrupted, and in the order they were sent [35]. In public data networks such as the Internet, data transfer reliability is mainly deteriorated by congestions. In such networks, achieving reliable data transfer is almost left to the endpoints, for example, by utilizing TCP transport layer protocol. Network-based approaches to improve the data transfer reliability start to appear in newer communication technologies such as the Universal Mobile Telecommunications System (UMTS) [44, 51]. These approaches are mainly based on differentiating and classifying users' traffic and associating it with certain forwarding treatments (e.g. resource allocation, prioritization, and packet error loss rate).

5 Reliable Communications Within CPSs

The reliability of a communication network becomes a crucial issue when considering CPSs. An unreliable communication might cause not only financial costs and service disruptions but also human fatalities. In fact, one of the key requirements to realize CPSs in critical infrastructure is to have reliable communication networks [37]. As mentioned earlier in this chapter, control loops in CPSs might be closed over networks. If the network is unreliable, then degradation or destabilization of the control loops are expected. Even though that several CPSs have requirements on the network performance; it is unlikely that such requirements can be guaranteed in an unreliable network. This is mainly due to severe and frequent disruptions of communications service that occur in unreliable networks [50].

To illustrate the importance of communication network reliability, we consider the following example. When a frequency event in a power grid occurs, such as a sudden loss of generation, the grid frequency response is divided into three phases [3]. One of these phases is the automatic generation control (AGC). In this phase, the grid utility operator sends power signals to the different power plants to adjust the level of generated power and restore the grid frequency to its nominal value. The time frame for the AGC to occur is between five and ten minutes. If we assumed that the communications service between the utility control center and the generation planets is unreliable, which might happen due to oversubscription, routing protocols failures, etc. In that case, the time frame to apply the AGC cannot be met with high probability. This, in turn, might prolong the duration of the frequency event and cause damage to customer appliances that are designed to work at a certain grid frequency. Another important issue to indicate here is the requirement on network performance. In this example, it is clear that the performance requirements on the communications service are low and a time delay of several seconds, for example, can be tolerated. On the other hand, communications service reliability is more critical for such application, especially with regard to its availability.

Recent works in the domain of CPSs [2, 6, 41] have also confirmed the need for reliable communications, first indicated by Lee [36]. In addition, several works considered communication reliability for CPSs. As an example, Faza et al. [17] and Singh and Sprintson [49] presented the impact of cyber part reliability including the communication network on the overall smart grid's reliability. In [1, 39, 57, 60], the authors considered improving communication reliability for CPSs by proposing new designs for the communication networks and protocols as well as new architectures for communication networks. Later in this chapter, approaches for improving communication reliability for CPSs using redundancy will also be presented.

Improving reliability of communication networks is not a new issue that arose in the domain of CPSs, rather a long-term challenge in different domains and mostly associated with public (general purpose) communication networks. In this context, several common mechanisms have been developed in order to improve the network and/or data transfer reliabilities which include redundancy, equipment durability, resource reservation, prioritization, and endpoints data transfer protocols. In the following section, we will present some existing approaches to improve communications reliability.

6 Approaches to Improve Communications Reliability

Existing and proposed approaches to improve communications reliability might utilize more than one of the mentioned mechanisms in the previous section (i.e. redundancy, resource reservation, prioritization, etc.). Approaches that use only resource reservation and prioritization are more suitable for public networks due to the lack of control over their topologies and equipment by end users. On the other hand, approaches considering network reliability and its improvement by, for example, introducing redundant equipment and links seem more suitable for private/industrial networks particularly during the design phase.

6.1 Network Provided QoS Provisioning

Recent efforts to improve the reliability of general purpose public networks resulted in the definition of different service classes within the standards of different communication technologies. These service classes provide different levels of communication service availability based on the granted resources and prioritization to each of them. Nevertheless, and due to the different communications media and targeted performance of each of these communication technologies, there is an inconsistency between the service classes, and/or targeted QoS that each service class should receive in these technologies. For example, the telecommunication standardization sector of the International Telecommunication Union (ITU-T) defines six QoS classes for the different types of traffic in IP networks [29] while the Long-Term Evolution (LTE) standard [52] defines nine fine-grained QoS classes. As a result, one of the main challenges here is how to utilize and/or maintain such network provided QoS in the existence of heterogeneous communication technologies.

6.2 Redundancy

Redundancy is one of the widely used means to provide fault-tolerance and network reliability [4]. It can be an architectural property of the network that provides multiple elements in the network in a way that if one element fails, then one of the other available elements will provide the service or function. Network redundancy can be introduced at different levels including the overall network, devices, links, systems, and applications [40]. The obvious main advantage of redundancy is mitigating the single point of failure situation that exists when the failure of a single network element can cause disruption of the service. Therefore, redundancy has been traditionally adopted in telephone networks to improve reliability and customer satisfaction [32]. In addition, most of the international ISPs, known as tier 1 ISPs, usually employ fault-tolerance techniques which are mainly based on hardening devices and introducing redundant components and links. In such ISPs, load balancing schemes might also be utilized in combination with redundancy to improve network performance and balance the utilization of primary and redundant links [14, 24].

Utilization of network redundancy has also been considered in some standards and existing implementations of CPSs where a high level of communication reliability is required. The requirements on the WAN described in the NASPInet specification project [25] (initiated by the North American Synchro-Phasor Initiative (NASPI)) clearly state that the private WAN should be able to provide redundant paths for data classes requiring high availability. IEC 62439 standard [34] is yet another example where network redundancy was exploited to improve communication reliability. IEC 62439 proposes a parallel redundancy protocol

(PRP) for local area networks with a requirement of two cloned (disjoint) but bridged networks. Here the MAC frames from the sender are duplicated and sent over the two interfaces connecting to the two LANs. Approaches to extended PRP to support IP networks were proposed in [45, 47]. Rentschler and Heine's approach was proposed to keep PRP information for duplicate identification included in the source MAC address while enabling PRP operation across routers. In contrast, Popovic et al. proposed a transport layer protocol with the requirement of disjoint (redundant) paths to improve reliability and fulfill the hard-delay constraints for smart grid applications within substations.

Data redundancy was also one of the considered approaches to improve communication reliability. This was achieved by means of information coding to provide packet redundancy (by generating a larger number of packets than the received number of application messages) in telesurgical robot systems [55]. Here, a minimum number of coded packets needs be received in order to retrieve all application messages.

6.3 New Network Generations

In this section, we will briefly introduce recent efforts in order to improve communications reliability in terms of data transfer reliability. SDN is one of these attempts that proposing new networking architecture [44]. Here, the core principle is to make the network more dynamic by decoupling the control plane of the network, including packet forwarding logic, from the network infrastructure. This is accomplished by logically centralizing the network control through software-based controllers in the control layer of the SDN. The SDN controllers also provide the application layer with an abstracted view of the network to look like a single logical programmable entity. This, in turn, provides unprecedented flexibility to network operators to adapt the network to their business needs. The proposed network architecture by SDNs target addressing the following points:

- The involved complexity to adapt existing networks to business needs, particularly, when implementing network-wide policies and large network expansions (which is mainly attributed to device-level tools for network management).
- Open interfaces provided by the vendors of network devices which increase the efforts to adapt new devices to business needs.

As a new network architecture with open issues such as the problem of how the application will communicate with the SDN control layer through the software-based controllers, there is a limited adoption for it in the meantime. Experience from standards related to networking such as IPv6 is not optimistic. Where after 15 years of being standardized, only a very small fraction of today Internet has migrated to IPv6 [20]. In addition, issues of how to charge for network resources reservation in public and commercial networks, the complexity of implementation in large-scale networks, and the need to upgrade existing infrastructures make it more appropriate for small or still in design phase networks.

Delay tolerant networks (DTNs) [16] are another relatively new networking paradigm to improve reliability when interconnecting different incompatible networks or when networks feature intermittent connectivity (e.g. mobile ad hoc networks (MANETs)). To accomplish such goals, DTNs use the store-and-forward scheme and translation between the different protocols adopted by the different networks. This is achieved using an overlayed transmission protocol, called the bundle protocol, on top of existing ones. However, the utilization of DTNs for CPSs seems not practical. From one hand, DTNs require modifications to existing routers in order to perform the persistent store of messages (one or more packets of an application). On the other hand, the use of store-and-forward scheme will increase the time delay experienced by packets traveling through the network. Moreover, complete or partial message store before forwarding it is not practical for nodes in core parts of large networks (e.g. routers), which handle large amounts of traffic.

7 CPS Communications Using the Internet

The Internet is a global data network of interconnected communication networks. It was defined by the U.S. Federal Networking Council resolution on October 24, 1995, to be a global information system that is logically linked together by a globally unique address space based on the Internet Protocol (IP). In the Internet, networks and routers that are under the control of a single administrative entity are usually referred to as an Autonomous System (AS). In some cases, two or more ASs might belong to the same administrative entity. Each AS is assigned a unique number, known as the AS number (ANS), and specific blocks of IP addresses. This allows the different ASs on the Internet to acquire a way to reach each other. Hence, the Internet can be considered as a system of interconnected ASs. As shown in Fig. 6, the architecture of the Internet can be considered to consist mainly of three levels or tiers. We first have the access networks (access ISPs) to connect end users to the Internet using a variety of wired and wireless technologies. Second, we have the regional ISPs that connect the different access ISPs on a regional level. Third, we have the tier 1 ISPs that connect to other tier 1 ISPs and, consequently, connect subscribed regional ISPs at different regions of the world. Each of these smaller networks might have different network service provider (NSP) policies and different link and physical layer technologies, but all are connected logically using the IP protocol.

The motivation to use the Internet for CPSs is more prominent when considering that some CPSs span large geographical areas where the control systems are located far away from each other. Examples of such CPSs are those in critical infrastructure including water and power utilities, agricultural production and transport systems, police and military security systems, and transportation systems (e.g. oil and gas

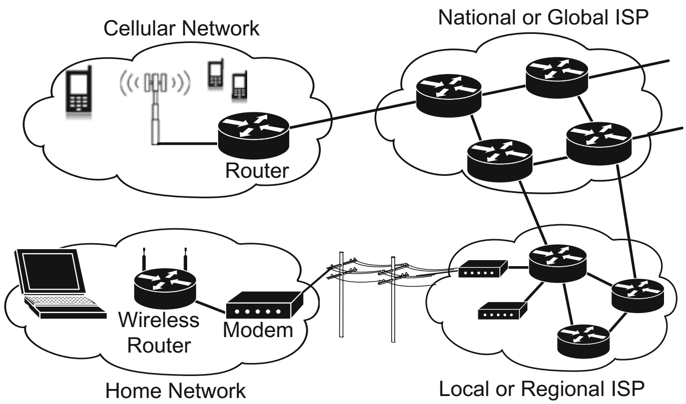


Fig. 6 Simplified topology of a very small part of Internet

transport pipeline systems, railways systems, shipping systems, etc.). As a detailed example, we consider smart grids. The U.S. Department of Energy (DOE) (2008) defines a smart grid as a widely distributed energy delivery network that features bi-directional flow of information and electricity and capable of monitoring everything from power plants in generation domain to individual appliances in distribution domain. The main target here is to achieve a near-instantaneous balance of power generation and consumption through the utilization of distributed computing and communications. In a smart grid, several control subsystems control and automate the different components including utility control center, power generation plants (e.g. fossil fuel power plants and wind farms), power transmission and distribution substations, and many other subsystems. For such systems, the Internet represents a very attractive solution to realize WAN communications as the use of dedicated communication networks is more expensive. Among the attractive features of Internet, we shortly describe a few here. The Internet offers almost worldwide connectivity with a wide range of wired and wireless access technologies that fulfill the different requirements of different geographical locations. The low cost of communications using the Internet is yet another important feature. Also, the Internet offers high flexibility to do changes later with regard to, for instance, the used access physical medium or the needed data rate without entailing high costs (e.g. by choosing another service provider).

On the other hand, The Internet consists mostly of shared general purpose public and commercial networks and characterized by offering the best-effort type of service with random time delay and packet loss [30]. Consequently, there is no upper bound on the time delay that the packet will experience when sent through the Internet. This is mainly attributed to the high heterogeneity of the smaller networks constituting it. The random nature of Internet time delay contradicts with the key requirement of the bounded time delay of many of the proposed CPS applications. Moreover, the heterogeneity of networks to access the Internet and connect its different parts and the high variety of their characteristics also render the Internet as an unreliable data network. This is mostly stemming from the nature of each of the individual networks where each has different reliability limitations. As mentioned before in this chapter, such limitations are attributed to failures, over-subscriptions, and other anomalies in the network. Therefore, the Internet inherently does not provide the needed service grantees for reliable operation of CPSs [12]. With this regards, the unreliable nature of commercial network services (which are part of the Internet) was presented in US. DOE [56] to indicate why most utility providers prefer to use their dedicated communication networks. The issues raised in US. DOE [56], included the insufficient backup power capabilities in commercial networks (with a high dependency on AC power from the power utility) and the inadequate priority of service provided. Utilities comments indicated that NSPs might not be able to maintain the required communications service availability due intermittent congestions and power disruptions. In addition, the unreliability of telephone and data networks in The USA and the challenges to providing reliable communications were also discussed by Snow [50]. Among the mentioned factors that contribute to reducing network reliability are complexity caused by concentrated infrastructure, the variety of used technologies and provided services, business revenue plans, market competition (which necessitate fast market entry), and rapid technological advances. All of these factors caused several service outages and severe service disruptions for the considered networks. For example, tracked outages affecting more than 30,000 service users for 30 min or more over a period of eight years were reported to occur 14 times every month approximately. In the examples provided in Snow [50], outages were usually occurring after certain events such as procedural errors and software and hardware upgrades. This in its turn might be attributed to non-highly qualified staff and difficulty to estimate all interactions in complex and large networks. Lastly, the lack of timing predictability in existing general purpose networks and the Internet and the need for reliable communication networks have been considered of key importance for CPSs in the literature (e.g. in [21, 36, 38]).

8 Communication Standards for CPSs

Previous work in estimating the traffic characteristics and communication requirements of CPSs and their associated applications revealed the wide spectrum of such requirements. It also indicated the difficulty involved in developing techniques for existing communication networks to fulfill the requirements of every single application. Moreover, there is a variety of technical approaches by the different vendors for automation, protection, and monitoring devices and need to enable interoperability and efficient integration of new systems and devices to existing ones. This motivated the development of different standards with this regards to provide a general basis to describe the communication requirements of CPS applications classes (e.g. protection applications). Another goal of these standards is to assist in the development of new communication technologies that take into consideration the communication requirements of CPSs. Consequently, efforts by international bodies to standardize requirements, services, and structures of communication in certain CPSs domains have been carried out. Power grids seem one of the earlier CPSs domains that witnessed a lot of research due to their expected economic and environmental impacts and also the development of some standards, such as IEEE C37.118 (IEEE 2006) and IEC 61850 (IEC 2004), that consider their communication requirements. As these standards will be the basis for the development of CPSs in the future, it is necessary to consider them in our investigation of the communications for CPSs. Here, we will cover only the IEC 61850 that represents a prominent standard for smart grids. The standard considers several aspects of the communication network for distributed energy resources (DER) such as the architecture of the network, communication services and their mapping, data transmission, and the communication requirements of the different performance classes. It also considers additional applications besides those considered in IEEE C37.118 (e.g. metering).

IEC 61850 specifies a communications model for the microcontrollers (referred to as Intelligent Electronic Device (IED) in the standard) connecting DER such as the photovoltaic systems and wind turbines to the communication network of the power grid. This enables the exchange of information between the DERs and the utility operator control center for monitoring and maintenance purposes and for consumed/produced energy charging. In addition, the standard defines 5 communication services for the different applications with different characteristics, which are: Abstract Communication Service Interface (ACSI), Generic Substation Status Event (GSSE), Generic Object Oriented Substation Event (GOOSE), Sampled Measured Value multicast (SMV), and Time Synchronization (TS) [58]. These communication services along with the communications stack of IEC 61850 are illustrated in Fig. 7. In addition, ACSI includes a number of client/server communication services which are defined in the IEC 61850-7-2 part of the standard. Examples of such client/server services include the *GetDirectory, GetDataObjectDefinition, Reporting, logging, GetDataValues* and *SetDataValues*. The *GetDirectory* service is used between IEDs to retrieve accessible objects (i.e. objects that describe the information generated and used by the different automation function within an IED) in each of them. The *GetDataObjectDefinition* service is used by clients to obtain the types of data for all accessible objects in an IED. The *Reporting* service is used to provide event-based data exchange between a server and client. The *logging* service is used to provide data archiving for clients. Finally, the *GetDataValues* and *SetDataValues* services are used to read and write the data objects values.

These client/server ACSI communication services are mapped to the Manufacturing Message Specification (MMS) protocol which in its turn is placed on top of the TCP/IP transport layer. Here, the MMS protocol specifies how the messages are structured to control and monitor the devices while the transmission of messages

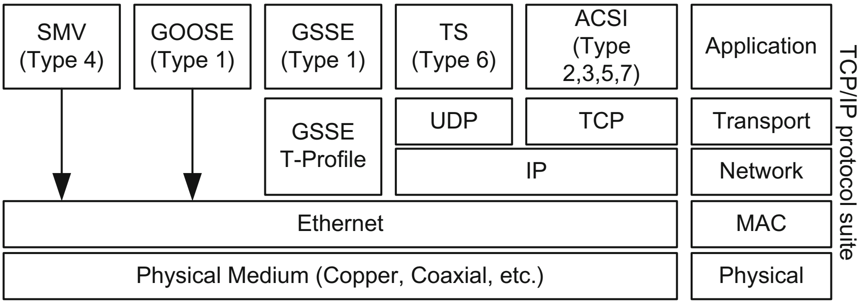


Fig. 7 The communications stack of IEC 61850

between the devices is handled by the TCP/IP protocol stack. Consequently, the client/server services can be realized using IP-based WANs for, for example, substation-to-substation communications. In the same way, the TS service uses the Standard Network Time Protocol (SNTP) that uses the UDP/IP protocols to synchronize the clocks of IEDs with a central timing entity (e.g. a time server). This service is used for applications with non-stringent requirements on maximum transfer time of messages but rather on accuracy. By contrast, the GOOSE, GSSE, and SMV communication services are time-critical and require specific mapping than that of ACSI and TS services. As illustrated in Fig. 7, The GOOSE and SMV services data are encapsulated directly in Ethernet PDUs while the GSSE service data are first mapped to the GSSE T-Profile PDUs. This is mainly to minimize protocol overhead and transfer time for these services as they are time critical. The GOOSE and GSSE communication services are used to convey status and events while the SMV communication service is used to convey sampled measured values from analog signals.

Furthermore, the standard defines seven types of messages with different transfer time requirements for the different communication services. As a result, applications utilizing the IEC 61850 communication services are classified by the standard based on the transfer time limits of these messages. This classification is shown in Table 1. As we can see from the table, GOOSE and GSSE communication services use the 1A messages with a requirement on maximum transfer time of 100 ms. ACSI services, on the other hand, might utilize messages of type 2, 5 or 7, based on the application, with requirements on maximum transfer time of 100 ms for type 2 and type 7 messages. SMV communication service uses type 4 messages due to their time critical nature with a very stringent requirement on the maximum transfer time to be less than or equal to 10 ms. Lastly, messages of type 5, 6A, and 6B do not have a specific bound on maximum transfer time and used for applications such as file transfer and time synchronization (for time synchronization the maximum transfer time is determined by the needed accuracy).

As we have mentioned earlier in this section, IEC 61850 specify a number of communication services that determine how the different components should interact with each other. In the case of the client/server ACSI services as an

Table 1 IEC 61850 message types and performance requirements

Message type	Application	Services	Transfer time required (ms)
1A	Fast message (trip)	GOOSE, GSSE	3–100
1B	Fast message (other)		20–100
2	Medium speed	ACSI	100
3	Low speed		500
4	Raw data	SMV	3–10
5	File transfer	ACSI	$\geq$ 1000
6A	Time synchronism messages A	TS	(Accuracy)
6B	Time synchronism messages B	TS	(Accuracy)
7	Command messages with access control (medium speed)	ACSI	100

example, the interactions follow a request-response pattern between the different components. However, different CPSs with different communication protocols might follow different communication patterns. The importance of such patterns for the development of future CPSs and the efforts carried out in this domain are detailed in the following section.

9 Communication Patterns for CPSs

A pattern is a generic, established, tested, and reliable concept for the solution of a problem. In the context of telecommunication, a communication pattern [48] is a commonly recurring interaction or task between the communicating parties. The main benefit in using patterns during solutions development is that existing concepts can be reused, and individual instances can be grouped into more generalized ones. In pattern-based development, matching patterns are chosen from a library, adapted, and combined. This process creates reliable solutions in a reasonable amount of time. The development effort can be focused on problems that are not solved yet. Christopher Alexanders in his book “The Timeless Way of Building” [5] pioneered the idea of patterns in the 1970s. While investigating the challenge of creating timeless buildings and towns, he recognized that different kinds of existing architectures implement similar solutions serving similar purposes. The reuse and rely on already established solutions have also been considered in the design of software and technical systems. Initial works proposing libraries of atomic patterns applicable in various phases of software development (e.g. architecture) were presented in [10, 19]. Each pattern is represented by a description of context, a problem, and a solution in order to solve the basic problems and introduce a common vocabulary. Several works in the literature highlighted the importance of communication patterns and extracted existing ones in certain CPS applications. We will list some of these works here for interested readers.

Brambilla et al. [8] investigated Service-oriented Architecture (SoA) protocols for business applications. They extracted asynchronous patterns such as callback, publish-subscribe, and polling and listed which interactions are present in the investigated protocols. They conclude that the callback pattern, also known as request-reply, is the one used by five of the six investigated protocols. Wu et al. [59] investigated the mutualities of MANETs, wireless sensor networks (WSNs), and CPSs and extracted three communication patterns for communication flows: Query-response (supported by MANET, WSN, and CPS), arbitrary communication (supported by MANET and CPS), and cross-domain communication (supported by CPS). Schoch et al. [48] analyzed envisioned applications for vehicular ad hoc networks (VANETs) and extracted a number of communication patterns where almost all considered VANETs applications were developed based on them. The extracted patterns, which target facilitating future development of communications systems for such networks, are beaconing, geobroadcast, unicast routing, advanced information dissemination, and information aggregation.

Henneke et al. [22] evaluated communication protocols in existing CPSs that are based on the concept of service-oriented architecture (SoA) and loosely coupled systems. The authors extracted common communication patterns by considering generic shared tasks and common implementations of relevant protocols. Among the considered protocols and middlewares for CPSs are OPC Unified Architecture (OPC-UA), web services, Universal Description, Discovery, and Integration (UDDI), and the Data Distribution Service (DDS). The work identified discovery, request-response, publish-subscribe as common patterns between these protocols and proposed two atomic communication patterns for future developments of CPS protocols, namely request-response and discovery. In the following, we briefly describe these three identified communication patterns.

9.1 Request-Response

In this pattern, client nodes initiate service requests (e.g. retrieve or modify data in a database, request calculation, etc.), and server nodes respond to those service requests. This pattern requires that clients are aware of service providers (server nodes). The request-response pattern is more suitable when the rate of a service call is limited, and the information content in the request is changeable.

9.2 Discovery

In this case, the endpoints providing services are dynamically selected rather than statistically configured and are searchable. This pattern requires that service providers provide Application Programming Interfaces (APIs) for their services. The pattern usually incorporates three entities, a service provider, a service consumer, and a directory of available services. Here, the service providers publish their services APIs to the service directory and the clients fetch the directory of needed services to find corresponding service providers to connect to them.

9.3 Publish-Subscribe

Here, the different nodes can be publishers, subscribers, or have both roles. Whenever a publisher node has a message, it publishes it to a corresponding topic to be automatically transferred to the other nodes subscribed to the topic. A consumer node, on the other hand, consumes the message content whenever it needs the content.

The above discussion clearly indicates the importance of patterns in systems and communication protocols development and the need for further work that focus explicitly on communication patterns for CPSs. By considering more CPSs and corresponding protocols to find more common communication patterns, it is expected that the effort to develop new protocols for CPSs will be smaller. This can be achieved when there is a large library of patterns available to be used to build the overall targeted communication interaction or at least most of it.

10 Conclusion

This chapter has described the aspect of communications for CPSs and the vital role of communication networks at all levels of CPSs from the field level up to the enterprise level. CPSs utilize different types of communication networks that include control networks as well as general purpose networks. For such networks, the reliability of communications is one of the important requirements that have been emphasized in the literature. This stems from the high reliability of operation expected from CPSs, especially when used in critical infrastructure. The high reliability of CPSs in these domains will ensure proper functioning of the society as well as the economy of nations. Even though that existing industrial networks forming the basic brick and nerve system of most existing CPSs, they are mostly local networks, require special devices and protocols, difficult to be integrated with widely used communication technologies, and do not fulfill future expectation of communication networks for CPSs (e.g. not confined in space). On the other hand, general purpose networks and the Internet mostly do not fulfill the requirements of communication reliability and timing predictability needed for many CPSs. In other words, existing industrial networks were designed for static distributed embedded systems with well-defined communication and control system requirements. The main focus of these networks is minimizing network latency. By contrast, public and commercial networks constituting Internet were designed for general purpose communications with a focus on throughput and scalability. Consequently, it becomes necessary to develop standards that consider the communications for CPSs and facilitate their real-world realization. The already carried out efforts in standardizing the communication network requirements of some CPSs, as in the case of the IEC 61850 for substation automation, represent a promising starting point in this domain. With the existence of standards such as IEC 61850, standardization bodies, such as the 3rd Generation Partnership Project (3GPP), can consider such communication requirements in standards of upcoming technologies. This will fill the gap between general purpose and industrial networks existing today and enables full utilization of CPSs potentials in our life. Previous work in other domains such as software design and technical systems indicated the importance of pattern-based development and motivated the investigation of communication patterns for CPSs. The obtained results in this area emphasize the existence of common communication patterns for CPSs where the identification of such common patterns will assist in the future development of communication protocols for CPSs.

References

1. Ahmadi H, Abdelzaher T (2009) An adaptive-reliability cyber-physical transport protocol for spatio-temporal data. In: 2009 30th IEEE real-time systems symposium. IEEE, pp 238–247
2. Ahmed SH, Kim G, Kim D (2013) Cyber physical system: architecture, applications and research challenges. In: 2013 IFIP wireless days (WD). IEEE, pp 1–5
3. Aho J, Buckspan A, Laks J et al (2012) A tutorial of wind turbine control for supporting grid frequency through active power control. In: 2012 American control conference (ACC). IEEE, pp 3120–3131
4. Alexander C (1979) The timeless way of building, vol 8. Oxford University Press, Oxford
5. Al-Kuwaiti M, Kyriakopoulos N, Hussein S (2009) A comparative analysis of network dependability, fault-tolerance, reliability, security, and survivability. IEEE Commun Surv Tutor 11:106–124. doi:10.1109/SURV.2009.090208
6. Babazadeh D, Chenine M, Nordström L (2013) Survey on the factors required in design of communication architecture for future DC grids. In: 2nd IFAC workshop on convergence of information technologies and control methods with power systems, ICPS 2013, Cluj-Napoca, Romania, pp. 58–63, 22 May–24 May 2013
7. Boswarthick D, Elloumi O, Hersent O (2012) M2M communications: a systems approach. Wiley, New York
8. Brambilla M, Guglielmetti G, Tziviskou C (2005) Asynchronous web services communication patterns in business protocols. In: Ngu AHH, Kitsuregawa M, Neuhold EJ, Chung J-Y, Sheng QZ (eds) Web information systems engineering—WISE 2005. Springer, Heidelberg, pp 435–442
9. Budka K, Deshpande J, Thottan M (2014) Communication networks for smart grids. Springer, London
10. Buschmann F, Meunier R, Rohnert H (1996) A system of patterns: pattern-oriented software architecture. Wiley, New York
11. Curtis K (2005) A DNP3 protocol primer. DNP User Group, pp. 1–8

12. Demir K, Germanus D, Suri N (2014) Robust and real-time communication on heterogeneous networks for smart distribution grid. In: 2014 IEEE international conference on smart grid communications (SmartGridComm). IEEE, pp 386–391
13. Dye M, McDonald R, Rufi A (2007) Network fundamentals, CCNA exploration companion guide. Cisco Press, USA
14. Elwalid A, Jin C, Low S, Widjaja I (2001) MATE: MPLS adaptive traffic engineering. In: Proceedings IEEE INFOCOM 2001, conference on computer communications. Twentieth annual joint conference of the IEEE computer and communications society (Cat. No. 01CH37213). IEEE, pp 1300–1309
15. ETR003 ETSI (1994) Network aspects (NA), general aspects of quality of service (QoS) and network performance (NP). Technical report
16. Fall K (2003) A delay-tolerant network architecture for challenged internets. In: Proceedings of the 2003 conference on applications, technologies, architectures, and protocols for computer communications—SIGCOMM 2003. ACM Press, New York, p 27
17. Faza A, Sedigh S, McMillin B (2010) Integrated cyber-physical fault injection for reliability analysis of the smart grid. In: Schoitsch E (ed) Proceedings of computer safety, reliability, and security: 29th international conference, SAFECOMP 2010, Vienna, Austria, 14–17 September 2010. Lecture notes in computer science, vol 6351. Springer, Berlin, pp 277–290
18. Galloway B, Hancke GP (2013) Introduction to industrial control networks. IEEE Commun Surv Tutorials 15:860–880. doi:10.1109/SURV.2012.071812.00124
19. Gamma E, Helm R, Johnson R, Vlissides J (1994) Design patterns: elements of reusable object-oriented software. Pearson Education, India
20. Haddadi H, Bonaventure O (eds) (2013) Recent advances in networking. ACM SIGCOMM, Brazil
21. Hauser CH, Bakken DE, Bose A (2005) A failure to communicate: next generation communication requirements, technologies, and architecture for the electric power grid. IEEE Power Energy Mag 3:47–55. doi:10.1109/MPAE.2005.1405870
22. Henneke D, Elattar M, Jasperneite J (2015) Communication patterns for cyber-physical systems. In: 2015 IEEE 20th conference on emerging technologies and factory automation (ETFA). IEEE, pp 1–4
23. Hespanha J, Naghshtabrizi P, Xu Y (2007) A survey of recent results in networked control systems. Proc IEEE 95:138
24. Hopps C (2000) Analysis of an equal-cost multi-path algorithm. IETF RFC 2992
25. Hu Y, Donnelly M, Helmer T, et al (2010) NASPInet specification—an important step toward its implementation. In: 2010 43rd Hawaii international conference on system sciences. IEEE, pp 1–9
26. IEEE (2012) IEEE standard for electric power systems communications—distributed network protocol (DNP3). IEEE Std 1815–2010
27. ITU (2014) The tactile internet. http://www.itu.int/en/ITU-T/techwatch/Pages/tactile-internet.aspx. Accessed 12 May 2016
28. ITU-T (1994) Terms and definitions related to quality of service and network performance including dependability. Recommendation E 800
29. ITU-T (2003) Network performance objectives for IP-based services. ITU-T Recommendation
30. Kamrani E (2012) Real-time internet-based teleoperation. Intell Control Autom 03:356–375. doi:10.4236/ica.2012.34041
31. Kamrani E, Momeni H, Sharafat A (2005) Modeling internet delay dynamics for teleoperation. In: Proceedings of 2005 IEEE conference on control applications, CCA 2005. IEEE, pp 1528–1533
32. Keshav S (1997) Engineering approach to computer networking: ATM networks, the internet, and the telephone network. Addison-Wesley Longman Publishing Co., Inc., USA
33. Khan R, Khan J (2013) A comprehensive review of the application characteristics and traffic requirements of a smart grid communications network. Comput Netw 57:825–845

34. Kirrmann H, Hansson M, Muri P (2007) IEC 62439 PRP: bumpless recovery for highly available, hard real-time industrial networks. In: 2007 IEEE conference on emerging technologies and factory automation (EFTA 2007). IEEE, pp 1396–1399
35. Kurose J, Ross K (2013) Computer networking: a top-down approach: international edition. Pearson Higher Education, USA
36. Lee EA (2006) Cyber-physical systems-are computing foundations adequate. In: Position paper for NSF workshop on cyber-physical systems: research motivation, techniques and roadmap. Citeseer
37. Lee EA (2008) Cyber physical systems: design challenges. In: 2008 11th IEEE international symposium on object and component-oriented real-time distributed computing (ISORC). IEEE, pp 363–369
38. Li H, Dimitrovski A, Bin SJ et al (2014) Communication infrastructure design in cyber physical systems with applications in smart grids: a hybrid system framework. IEEE Commun Surv Tutor 16:1689–1708. doi:10.1109/SURV.2014.052914.00130
39. Lin S-C, Gu L, Chen K-C (2015) Statistical dissemination control in large machine-to-machine communication networks. IEEE Trans Wirel Commun 14:1897–1910. doi:10.1109/TWC.2014.2376952
40. Liotine M (2003) Mission-critical network planning. Artech House, USA
41. Marashi K, Sarvestani SS (2014) Towards comprehensive modeling of reliability for smart grids: requirements and challenges. In: 2014 IEEE 15th international symposium high-assurance systems engineering (HASE), pp 105–112. doi:10.1109/HASE.2014.23
42. McCabe JD (1998) Practical computer network analysis and design. Morgan Kaufmann Publishers Inc., USA
43. NGMN Alliance (2015) NGMN 5G white paper. http://ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf. Accessed 12 May 2016
44. Open Networking Foundation (2012) Software-defined networking: the new norm for networks. ONF white paper
45. Popovic M, Mohiuddin M, Tomozei D-C, Le Boudec J-Y (2015) iPRP: parallel redundancy protocol for IP networks. In: 2015 IEEE world conference on factory communication systems (WFCS). IEEE, pp 1–4
46. Radatz J, Geraci A, Katki F (1990) IEEE standard glossary of software engineering terminology. IEEE Std 610121990:3
47. Rentschler M, Heine H (2013) The parallel redundancy protocol for industrial IP networks. In: 2013 IEEE international conference on industrial technology (ICIT). IEEE, pp 1404–1409
48. Schoch E, Kargl F, Weber M (2008) Communication patterns in VANETs. IEEE Commun Mag 46:119–125. doi:10.1109/MCOM.2008.4689254
49. Singh C, Sprintson A (2010) Reliability assurance of cyber-physical power systems. In: IEEE PES general meeting. IEEE, pp 1–6
50. Snow AP (2001) Network reliability: the concurrent challenges of innovation, competition, and complexity. IEEE Trans Reliab 50:38–40. doi:10.1109/24.935015
51. The 3rd Generation Partnership Project (3GPP) (1999) 3GPP release 99. http://www.3gpp.org/. Accessed 10 May 2016
52. The 3rd Generation Partnership Project (3GPP) (2008) Long-term evolution (LTE). http://www.3gpp.org/specifications/releases/72-release-8. Accessed 10 May 2016
53. The technology network: intelligent technical systems Ostwestfalen-Lippe it's OWL (2014) Towards industrial 4.0: solutions from the leading-edge cluster it's OWL. http://www.its-owl.de/fileadmin/PDF/Industrie_4.0/Auf_dem_Weg_zu_Industrie_4.0_-_Loesungen_aus_dem_Spitzencluster_its_OWL_RGB.pdf. Accessed 12 May 2016
54. Tipsuwan Y, Chow M (2003) Control methodologies in networked control systems. Control Eng Pract 11:1099–1111
55. Tozal ME, Al-Shaer E, Sarac K, Thuraisingham B (2011) On secure and resilient telesurgery communications over unreliable networks. In: 2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS). IEEE, pp 714–719

56. US. DOE (2010) Communications requirements of smart grid technologies. US Department Energy, Technical report 2010
57. Weiner M, Jorgovanovic M, Sahai A, Nikolie B (2014) Design of a low-latency, high-reliability wireless communication system for control applications. In: 2014 IEEE international conference on communications (ICC). IEEE, pp 3829–3835
58. Wilamowski B, Irwin J (2011) Industrial communication systems. CRC Press, Boca Raton
59. Wu F, Kao Y, Tseng Y (2011) From wireless sensor networks towards cyber physical systems. Pervasive Mob Comput 7:397–413
60. Wu X, Dong Y, Ge Y, Zhao H (2013) A high reliable communication technology in electric vehicle charging station. In: 2013 IEEE seventh international conference on software security and reliability companion. IEEE, pp 198–203

Part V
Artificial Intelligence and Data Analytics for Manufacturing

Application of CPS in Machine Tools

Christoph Berger, Juliane Nägele, Benny Drescher and Gunther Reinhart

1 Motivation for CPS in the Manufacturing Environment

Industrial production is facing major challenges with regard to flexibility and productivity during the manufacturing process. Due to the dynamization of product lifecycles, the number of product variants manufactured in the same production environment is continuously increasing. In addition, a high productivity rate is required in order to remain competitive as the competition in the production industry increases. In order to deal with these challenges, the need for efficiency, flexibility and responsiveness arises in the area of manufacturing processes [1]. The objective here is intelligent linked-up production, which can be defined as the manufacturing of products on an integrated information technology basis via the usage of digital technologies with respect to products, processes and resources in line with the production development process.

A decentralized intelligence, built up by sensors and actuators, combined via the appropriate infrastructure serves as a solution to handle the tradeoff between productivity and flexibility. This leads to an increasing impact of the application of cyber-physical systems (CPS) and information and communication technologies (ICT) in the manufacturing systems [5].

C. Berger (✉) · J. Nägele · B. Drescher · G. Reinhart
Fraunhofer IGCV, Provinostr. 52, 86153 Augsburg, Germany
e-mail: Christoph.berger@igcv.fraunhofer.de

J. Nägele
e-mail: Juliane.Naegele@iwu.fraunhofer.de

B. Drescher
e-mail: Benny.Drescher@iwu.fraunhofer.de

G. Reinhart
e-mail: Gunther.Reinhart@iwb.tum.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_14

This article focuses on the area of machining, in particular on the processes of turning, milling and drilling. Current information about the state of the process or the product quality is usually not available. In view of those circumstances, the parameters are usually manually set by the operator [27].

Since the parameters are set static, a high safety coefficient is chosen to ensure a reproducible result considering the changing conditions of the environment. Due to the required resilience of the system, the consequence is a lack of performance and productivity [9]. Implementing CPS in the machining process allows dynamic and adaptive parametrization, as the integration of the sensor and actuator enables direct intervention in the process.

To analyze the potential of CPS in the machining environment from an industrial standpoint, a survey of 30 companies from different industrial sectors was conducted by the Fraunhofer IGCV. The focus of the survey was on the monitoring and control systems used in the machining processes. The objective is the identification of deficiencies which can be resolved by integrating CPS in the machine tool. In order to evaluate the functionalities which have to be considered in the development of such CPS, the main optimization potentials in the machining processes are identified in the survey, thus allowing process conditions and process failures to be distinguished. Considering the process conditions, machine productivity and size accuracy were named as the main areas which should be monitored and optimized in the machining processes. Regarding the process failures, the importance for the operator, the frequency of occurrence, and the risk of subsequent damage were evaluated, and tool breakage and tool wear were indicated as the most important sources of failure. Accordingly, these four identified fields form the influencing variables for the development process.

Besides the high impact of intelligence and mechatronics in the machining centers itself, the tools and clamping devices used are rather mechanical. In terms of decentralization, the application of information technology and embedded devices on the tool level provides a high potential [21]. However, the development of the CPS as a standalone solution is not able to fully utilize the potential to the maximum. Integration into an appropriate IT infrastructure, consideration of the real-time requirement, the requirements of the machining process and the production environment is necessary [5].

This article focuses on the application of CPS in the machining environment and the necessary manufacturing environment. To provide a comprehensive overview about the different aspects of the implementation, development and the diversity of the use cases, the state of the art regarding the characteristics, classification and requirements is explained in Sect. 2. Due to the different requirements, the communication structure of CPS can vary in each use case. A general approach to designing the information distribution in the manufacturing process is described in Sect. 3. Section 4 presents two solutions in two different areas of machining according to the defined approach of the information distribution. An evaluation and classification of the characteristics, classification and requirements of the presented use cases is provided in Sect. 5.

2 State of the Art—Literature Review

2.1 Characteristics of Cyber-Physical Systems in Manufacturing

Cyber-physical systems are one of the main enablers for flexibility and productivity in manufacturing processes in the future [10]. They consist of embedded systems with the ability to communicate, preferably via Internet technologies. These special types of embedded systems, based on powerful software systems, enable integration in digital networks and create completely new functionalities. Because of their digital representation, these systems are part of cyberspace. They are more than just an interface due to their ability to represent relevant knowledge about the physical reality and their autonomous computing capacity for analysis and interpretation of the data. A dominant characteristic of cyber-physical systems is their potential to interact with other CPS beyond their system boundaries. These systems are able to link the diverse subsystems. Therefore, the solutions must be designed for cross-linking [15]. Today, solutions with such networking abilities are not available. Without a networking standard, the possibilities for cyber-physical systems are limited. Cyber-physical systems enable completely new system functionalities and applications in manufacturing companies, such as condition-based maintenance and process regulation. An application example is the integration of new functions, which leads to the multi-functionality of the systems. CPS are able to collect data from the real applications, process them in complex algorithms and transfer the results to further embedded systems and to large central computing facilities. These structures may result in a high degree of flexibility [4].

The current definitions and literature show the properties of CPS and their respective environment. To this end, the essential characteristics from the publications and applications are summarized in the following [7, 23, 24, 28].

- The basic core of a CPS is an electronic system. The functions are largely determined by the software functions and not by the hardware.
- An exchange of information with other devices is not limited to CPS. Communication with other participants on the shop floor or in appropriate top-level systems, such as a programmable logic controller (PLC) or a ME system for example, should be taken into account.
- There are no standard communication interfaces postulated in the literature. It can be seen that open software standards are preferred for market penetration, IT security and cost reasons.
- The interaction with the environment is not limited to the exchange of information via a network. It is common to integrate sensors and actuators directly in

the CPS. By implementing that targeted measure, decentralized control and regulation for external influences can be established.
- One important aspect of cyber-physical systems is their self-description, i.e. the knowledge about their own status, possible reference functionality and the internal data.

The aforementioned properties are dependent on the application and technology in the manufacturing process. To call a device a CPS does not mean it necessarily has to fulfill all of the above features.

2.2 *Classification of Intelligent Objects in the Machining Process*

Based on the defined characteristics, there is still a wide spectrum of possible specifications for CPS. Besides this definition, the terms "intelligent product" or "intelligent object" are commonly used within the context of industrial production. Therefore, the following section introduces several approaches to defining an intelligent product. An analysis of the application of the presented classifications considering the machining processes will then be conducted.

The Oxford Dictionary defines an intelligent device as having the ability "to vary its state or action in response to varying situations, varying requirements, and past experience." This definition serves as a basis for the following approaches and will be detailed in the reviewed literature.

Kiritsis [8] describe four levels of intelligence that a device or a product can apply. He discusses the augmentation from the physical product without any interaction via a simple interaction to a product containing embedded information devices. Zbib et al. [33] define four similar classes with matched functionalities of sensor actuation, processing, communication and memorization. Starting from identification capabilities up to memory functionalities and decision-making functionalities, the highest level contains a product incorporating all functionalities with interaction capabilities. Lee et al. [11] describe a five-level guideline for the development of CPS. This extends from the connection level of the device, where data acquisition can be achieved, via several steps up to the configuration level, where actions can be performed.

In contrast to the definition of the intelligent product via several levels, [12] describe the devices analyzing their features. They identify five features which should be incorporated in an intelligent product: unique identification, capability to communicate with the environment, the ability to store and retain data, use of a language to display properties and requirements, and the capability to participate in the decision-making process about itself. Wong et al. [32] utilize McFarlane as a basis and extend the approach, defining two levels of product intelligence. The first level comprises the first three features stated by McFarlane et al. [12] and is defined as an information-oriented product. The second level incorporates all features and is

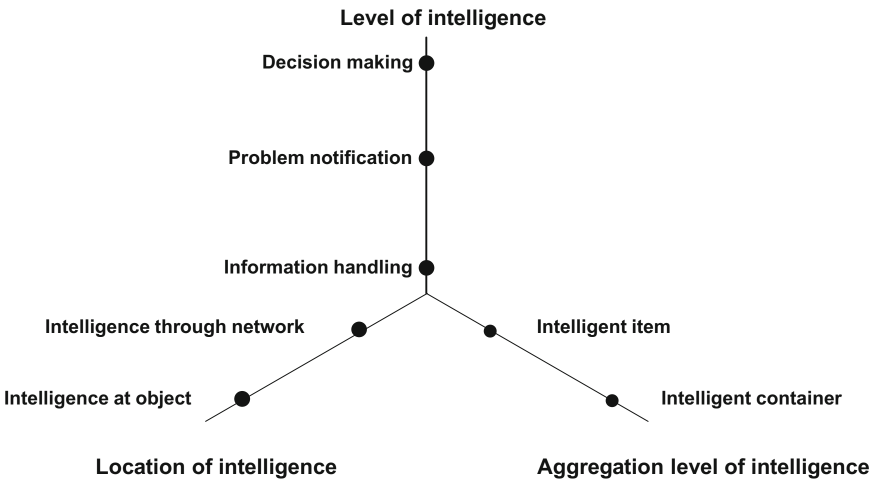


Fig. 1 Classification model of Meyer et al. [10]

defined as a decision-oriented product. [26] focuses on the decision-oriented products and selects the feature-oriented definition. The continuous monitoring, reaction and adaption to external conditions, the maintenance of optimal performance, and the active communication with other systems form the features which describe the intelligent products in this context. Meyer et al. [13] choose a method to combine the introduced approaches. The three axes in Fig. 1 with different dimensions are defined: level of intelligence, location of intelligence and aggregation level of intelligence.

The level of intelligence comprises three categories, from information handling and problem notification up to decision making. The dimension of the aggregation level opens a new perspective on product intelligence. A differentiation between an intelligent item (knowledge only about the object itself) and an intelligent container (knowledge about the included components) is represented in this axis. The third dimension comprises the location of the intelligence and distinguishes between the intelligence through the network or at the object [13].

To summarize, there are several existing approaches to defining an intelligent product. Intelligent products can be defined descriptively, by features, or include several levels of intelligence. All of those approaches focus on the areas of identification, decentralized data processing, and communication. From the manufacturing point of view, the definitions are rather general. Two open topics can be defined in order to apply a classification of the objects in the machining environment. In the first step, the term "product" or "object" has to be defined with respect to context. Following that, the categories are concretized with regard to the requirements of the manufacturing industry/machining processes.

As already stated, the production environment is characterized by a wide variety of equipment, resources, products, objects and devices. Different requirements concerning communication interfaces and integration are therefore dependent on the nature and configuration of the device in question [30]. In order to define the specific product precisely, a classification from the guidelines of simulation in the context of production systems has been used. The guideline differentiates between “building blocks” with a structural logic as an active element, and “objects”, which are defined as passive elements without their own internal logic [17].

Since the focus of this paper is on components of machine tools, the description of the passive elements is most suitable. In this context, an object is defined as a part of the machine tool which, in its initial state, does not have its own internal structural logic.

Furthermore, the definition of the intelligent product has to be extended and adapted to the manufacturing industry. Given that [13] consolidate several approaches, incorporating the capability to describe further dimensions to indicate the variety of the production environment, their approach is chosen as basis for the classification. Therefore, the axes are examined in further detail.

The dimension of the “aggregation level of intelligence” is indispensable in supply chain processes, but has no prior significance regarding machine components in the manufacturing environment. Since vertical integration forms one major focus in the application of CPS in the industrial production, the communication capabilities have to be taken into account. The automation pyramid defines possible levels concerning vertical communication. Therefore, the classification scheme is adapted in this dimension and three levels of the automation pyramid are applied. Beyond that, the “location of intelligence” axis represents possible characteristics of intelligent objects in the machining area and can be resumed without needing any adaptation. The third dimension, “level of intelligence,” is necessary to describe a device in the manufacturing environment. Since sensor and actuator integration is significantly involved in the evolvement of manufacturing objects, the “problem notification” classification is divided into passive and active problem notification. With this adaption, the differentiation can be displayed in the classification scheme. The adapted classification model is shown in Fig. 2.

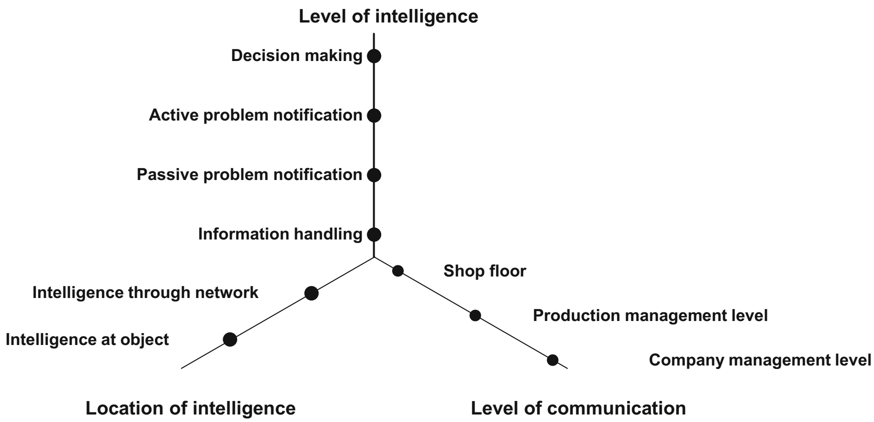


Fig. 2 Classification model for the manufacturing environment (based on Meyer et al. [13])

The classification model supports the evaluation of the devices considered. Since the requirements can be set by the production environment, the CPS needs to be developed to fit the appropriate level on the provided framework. After classification of the remaining components of the manufacturing environment, the interfaces can be designed appropriately. The model will be applied to the developed solutions, which is presented in Sect. 5.

2.3 *Definition of "Real Time" in the Context of Manufacturing*

To integrate intelligent objects/CPS in the industrial production environment, the requirements in the different production levels have to be fulfilled. Today, most production sites maintain roughly three levels, from the shop floor to the production (Manufacturing Execution Systems (MES)) up to the company management (Enterprise Resource Planning (ERP)) level. One reference architecture used to describe this structure is the information pyramid of automation [20]. The various tasks at each level lead to the different perception of the term "real time".

Real time is defined by the DIN 44300 standard as the process of a computing system in which programs for processing accruing data are constantly ready. The term "ready" indicates that the processing results are available within a predetermined period (DIN 44300-9). A real-time system is therefore a system which is able to process a specific task within a specific time window. In this case, it is not crucial to finish a task as soon as possible, it is more important to end at a fixed predefined time constraint. The execution of a task is fulfilled in time if the corresponding time requirement has been complied with. In addition, a distinction is made between hard and soft real-time requirements, seen in Fig. 3.

In a periphery with a hard real-time condition (A) an unmet target date (d) leads to an unacceptable error, see Fig. 3. With respect to the production, hard real-time conditions, such as turning, can be found on the shop floor level [34]. The sum of the start date (r) and the delta e (Δ_e) must always be less than the target date (d) in order to meet the hard real-time requirements, as shown in Formula 1.

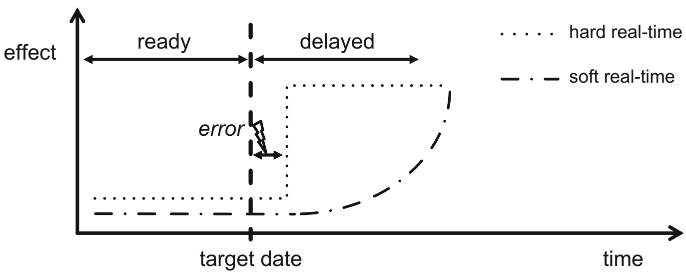


Fig. 3 Illustration of the hard and soft real-time requirement [34]

$$\text{Formula:} \qquad A \equiv r + \Delta_e \leq d \tag{1}$$

IEC 61784-2 (DIN EN 61784-3) proposes three real-time classifications, especially for the shop floor level. It refers to the maximum response time and also explicitly to suitable automation functions. For each real-time classification, Formula 2 has to be fulfilled. The hard real-time condition is characterized by the fact that the time condition A necessary at favorable conditions B, i.e. with probability (P) 1, is to be complied with.

$$\text{Formula:} \qquad P(A|B) \leq 1 \tag{2}$$

Soft real-time conditions allow the exceedance of a time limit as the target date (d). The system tolerates this condition and will continue to function, see Fig. 3. The longer the time limit is exceeded, the greater the negative impact on the system. In the production, we find some soft real-time conditions. One example is when a user starts a process and there is a delay between the user start time and the process start time.

In addition to the described timeliness, a real-time system must also have the property of predictability and determinism. Therefore, with regard to all information handling, we need a definition of the real-time requirement. That includes the processing time of the task until it receives the forwarded information from other systems.

Furthermore, in this article, the expected real-time conditions (A) will be indicated with the indices hard (h), soft (s), and the target date in milliseconds (ms), see Formula 3. Thus, the conditions are identifiable and verifiable.

$$\text{Formula:} \qquad A_{h,300} \tag{3}$$

This method is one step in creating transparency and determinism in the information flow in the manufacturing process.

2.4 Derivation of Requirements

Based on the literature review, concrete requirements are derived in order to establish a proven basis for the development of the CPS as part of the machine tools. The following functional requirements can be stated:

Characteristics:

- An embedded system must determine the information flow and provides an essential function in the process.
- Either horizontal or vertical communication needs to be implemented with the CPS.

- Decentralized control of sensors and actuators must be implemented.
- Self-description as knowledge about its own status, possible reference functionality and about the internal data must be incorporated in the CPS.

Classification scheme:

- All three dimensions of the classification model must be applicable to the CPS.

Real-time characteristics:

- Adequate communication technologies must be chosen in accordance with the real-time requirements of the process.

These requirements must be fulfilled by the solutions presented. To provide an overview of the implemented information and communication technologies, the approach to information distribution is presented in the following section.

3 Approach to Information Distribution

Information technology enables modern production systems to react to the requirements of the market, to operate in the network and to exchange information in real time. Integration of existing and new IT systems is very important in order to continuously support business processes.

Currently, we primarily find cyber-physical systems on the shop floor level. Nevertheless, the interfaces also allow an exchange of information with higher level systems such as ERP and MES. Figure 4 shows a classic automation pyramid with a 90° counterclockwise rotation. Current research topics such as "Internet of Things" change this hierarchical structure. These changes appear clearly in the information flow between the levels. In the future, an increase in the direct exchange of data between systems such as Manufacturing Execution Systems (MES), Enterprise Resource Management (ERP) and external services is to be expected (DIN SPEC 91329). Accordingly, the requirements regarding the communications and statements about the real-time element are increasingly becoming an important aspect in manufacturing. Consequently, an approach to the integration of CPS in the information architecture is presented by taking into account real-time requirements in the manufacturing environment.

The bases for this analysis are the information sources and sinks and the determination of the time requirement. The consideration of the information flow is not limited to the current and future approaches to the communications architecture. Furthermore it also reflects aspects of control engineering [17].

One of the defining characteristics of the automation pyramid is the chronological grades between levels such as shop floor and production management. The next step is to convert the chronological grades into a structure of the regulation

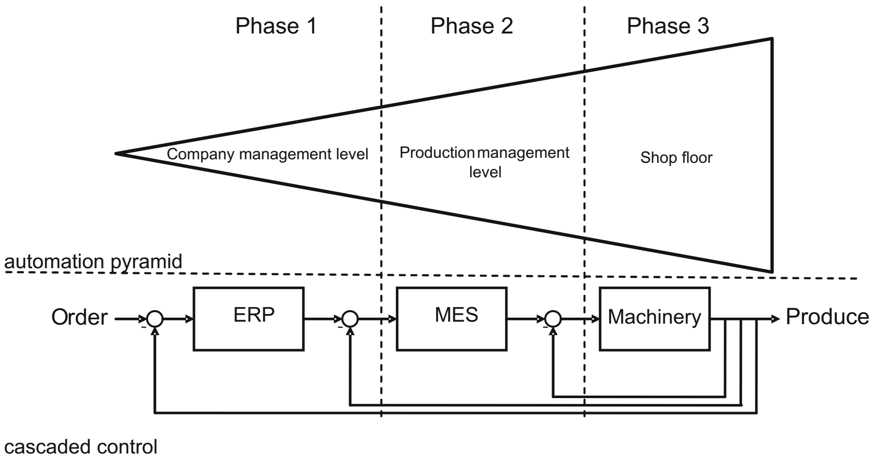


Fig. 4 Automation pyramid in comparison with cascaded control

technology it will represent as a cascaded control system, see Fig. 4 [12], [18], [19], [31].

The usage of a cascaded control structure offers several advantages. Disturbances caused on the shop floor are not immediately reflected in other levels. It is possible to dampen or avoid a bullwhip effect in the manufacturing. Secondly, the necessary performance can be determined accurately regarding the communication and the processing speed of the information.

Figure 4 shows the comparison of a three-level automation pyramid with a corresponding cascaded control system. The main topic in the first section (Phase 1) is concerned with tasks related to operations planning, order processing and long-term analysis of the production instead. This level can use internal and external resources to solve tasks by utilizing a cloud or a web-based service. The functionality on the second level includes detailed production planning, production data acquisition, key performance identification (KPI) and quality management. For this functionality, it is necessary to gather information about the first level and shop floor respectively. One of the requirements for controlling a production process in time is real-time decisions and another is high-quality information from the manufacturing about production events. The duties of the shop floor level mainly consist in the control, monitoring and regulation of plant and machinery. The tasks of the machine control unit include the production of the desired quantity of products, control of the drive units and execution of manufacturing steps. These specifications can be derived from an ERP or MES [2, 21].

Based on the comparison of the two concepts, as outlined in Fig. 4, and the well-known design rules for cascaded control technology, it is possible to create guidelines for the information flow in the production. These recommendations consider the possibility to integrate CPS into the common production environment. The following four guidelines serve to facilitate the chronological sequence in the different levels of manufacturing with CPS. These requirements combine the experiences from cascaded control technology, automation technology in manufacturing, aspects of real-time, and the characteristics of cyber-physical systems.

- The response times of a bad case scenario in and between the manufacturing levels need to be defined.
- The reaction time between two levels should be in the ratio of 1:10 or greater.
- For the communication speed between the levels, the following agreements must be maintained Phase 1 > Phase 2 > Phase 3.
- The real-time capability of the hardware must be verified.

The implementation of the above guidelines will be considered and addressed in the next section. For this purpose, the following notation for the information flow has been chosen, see Formula 4.

$$\text{Formula:} \quad \text{Source to Sink: } A_{H100} \rightarrow A_{H100} \rightarrow A_{H1000} \rightarrow A_{S2000}\text{: } A_{Sum} 3200\,\text{s} \tag{4}$$

Here, the information route from the source to the sink is observed in order to document real-time requirements for each connection. If any soft real-time requirements exist in the information flow, the sum will be marked accordingly with "S". A hard real-time requirement is no longer possible. The sum of the times is shown in milliseconds and equal to the maximum expected response time. This time can be used as a basis for the following processes in the manufacturing.

4 Solutions in the Area of Machining

Two examples from research projects are presented in the following section. The first example involves a chuck for a turning machine. In the second example, the development of a milling tool is described. Both examples are in their initial state without any internal structural logic. The evolvement of the examples is described in the three defined phases of the cascaded control systems as stated in Sect. 3. The requirements derived from the literature review will be validated following the presentation of the solutions in Sect. 5.

4.1 Intelligent Chuck for Turning Machine

The following use cases, shown in Fig. 5, for a CPS are based on the intelligent chuck for a turning machine, developed as part of the CyProS research project [6], [16].

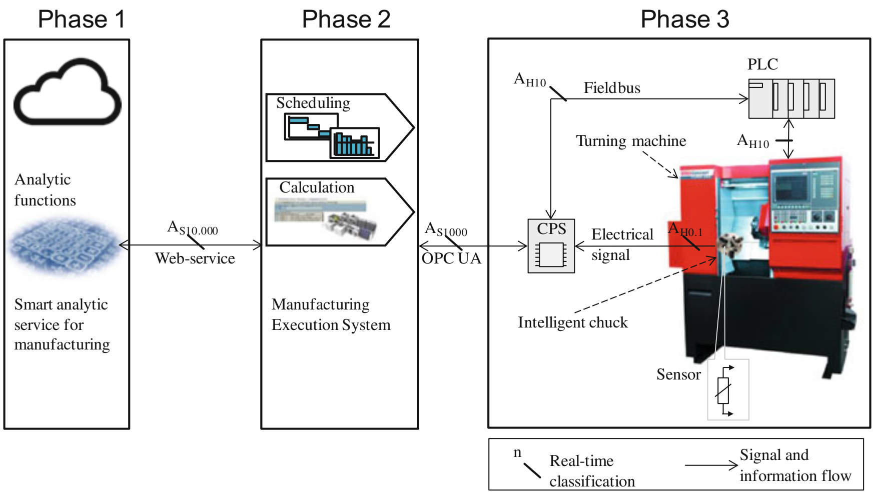


Fig. 5 CPS for a turning machine

4.1.1 Phase I: Company Management Level—Analysis of the Turning Process

At the corporate management level, we find some tasks upstream and downstream of the value-adding turning process. To link those functionalities to the actual turning process, it is possible to use computer resources in the companies as well external services, for example in a cloud. The following two examples show the possibilities for the turning process. In order to be able to use the abilities of an intelligent clamping tool, it is imperative to calculate the maximum clamping force for each assembly element. To this end, the information in the sales order about the material and the quality requirements can be used. This information will be synchronized with the machine status to calculate the parameters for the clamping force. After the turning process, all of the information from the intelligent chuck, the turning machine and operator can be analyzed. Thus, the acting forces on the assembly element will be evaluated by the motor currents from the turning machine and other signals through the use of special algorithms. If the analysis yields a noticeable result, the assembly elements are reviewed and it is possible to avoid quality deficits during the further production process.

4.1.2 Phase II: Production Management Level—Scheduling and Situational Production Control

The ME system in this application has a variety of functions. Some of these functions will now be explained. The first feature of the ME system is the selection of the correct CNC program according to the upcoming assembly element. Thus a high degree of transparency about the current process is created in the production, which reduces the possibility of errors. Furthermore, the actual force and pre-positioning values of the maximal force, the position of the chucks and the operating mode of the turning machine are transmitted to the turning machine. A complex and situation-dependent function of the ME system is the calculation of linearization values for the strain gauges. The reason is that strain gauge sensors have a curved characteristic. A linear relationship between the non-electric force and the output signal is useful. For this reason, curves must be linearized. This linearization happens either for a small range with accuracy or for a wide range with imprecision. This characteristic limits the effective measuring range. The different processes such as threading, grooving or knurling cause large fluctuations in the measurement range. With the communication to the PLC, the turning machine and the ME system information can be provided due to the expected force. With the expected load in the turning process, a custom linearization of the input signal can be carried out. The result is an accurate measurement from the strain gauges.

4.1.3 Phase III: Shop Floor—Development of a Chuck Control

This intelligent chuck controls and regulates the clamping force on the basis of sensor data, measured directly on the clamped part. In the horizontal automation hierarchy, the chuck is able to communicate with the PLC. On this basis, a requirement-ability comparison between product and chuck can be implemented. As a result, setup times can be reduced. Furthermore, this self-regulated turning process allows higher product quality and minimum component violation. In the vertical hierarchy, the intelligent chuck is fully integrated into ME system through the use of OPC UA (Object Linking Embedding for Process Control Unified Architecture) and Ethernet. The ME system thus represents the centralized information platform that enables situational control decisions and order scheduling. With the integration of a CPS, a situational production control is more accurate. A service-oriented sequencer communicates with the ME system via a web service and calculates the ideal order sequence depending on the queuing parts and the company-specific, prioritized logistics objectives, e.g. low throughput times. The prerequisite for utilizing these advantages is a robust and reliable measuring system.

In many cases, current systems for production planning and control cannot react to the actual situation on the shop floor. Most production planning decisions are based on historical data provided by production machines and usually inefficient operating data. A CPS-based production environment has the potential to collect status information from an intelligent sensor network. An example is the quality control of workpieces with an intelligent chuck, shown in Fig. 5. At present, quality criteria can only be applied in an additional workstation. By increasing the transparency about an intelligent chuck, additional information and relevant data are necessary (e.g. turning forces). In a potential overrun, the processing can be stopped and the next processing steps rescheduled. With the help of this procedure, ad hoc situation-based decisions are possible and enable manufacturing companies to ensure their production quality.

To utilize the advantages of an intelligent chuck, is it necessary to respect the real-time requirements concerning the signals and data flows, as explained in Sect. 3. The requirement regarding the real-time is indicated at the indices in Fig. 5. In phases 1 and 2, no hard real-time requirement is necessary. Here, a soft real time with the corresponding time requirements is adequate. In phase 3, hard real-time requirements are needed. A violation of these real-time requirements can cause damage to the machine or material.

To display the real-time calculation, the final analysis in an external service (Sect. 4.1.1) is used to calculate the "source to sink" time according to Formula 3. The information flow involves transmission paths with a soft real-time requirement. In accordance with the convention adopted in Sect. 3, the sum of the real-time requirements can only fulfill the soft requirement. The following calculation shows the compiled time. The following formula shows the real-time requirements for the pre-quality analyst function.

$$\text{Formula:}\quad \text{Source to Sink: } A_{H0,1} \rightarrow A_{H10} \rightarrow A_{H10} \rightarrow A_{H1000} \rightarrow A_{S10.000}: A_{Sum} 11.120, 1\ \text{s} \tag{5}$$

From this, it can be concluded that the necessary information is provided to the pre-quality analyst function in approximately 11.1 s.

4.2 *Intelligent Milling Tool*

Besides the clamping devices, the tools themselves have a major impact on the process and the product quality. Regarding the identified process conditions and failures which should be optimized, the development of an intelligent tool serves as an enabler for several functions. In the "BaZMod" research project, funded by the German Federal Ministry of Education and Research (BMBF), the objective is a definition of a standard for data and energy gateways inside the machine tool. This enables the application of mechatronic modules or CPS in the machining processes, see Fig. 6. The focus is on the development of the appropriate intelligent tools, including the IT infrastructure and the closed-loop control. Since the IT infrastructure is relevant throughout all phases, the technical realization will be described in the following.

The IT infrastructure is represented by various smart devices (e.g. intelligent tools) that can simultaneously connect to a company-specific gateway. The connection uses a variety of protocols. In our use case, WiFi, Bluetooth and OPC UA are used with the specific smart device. Apart from the intelligent tool, a Wifi simulator in Java and a Bluetooth simulator in Python were programmed and tested within the infrastructure. The gateway is implemented within the company and represents a tunnel to the Internet in general and to cloud services in specific.

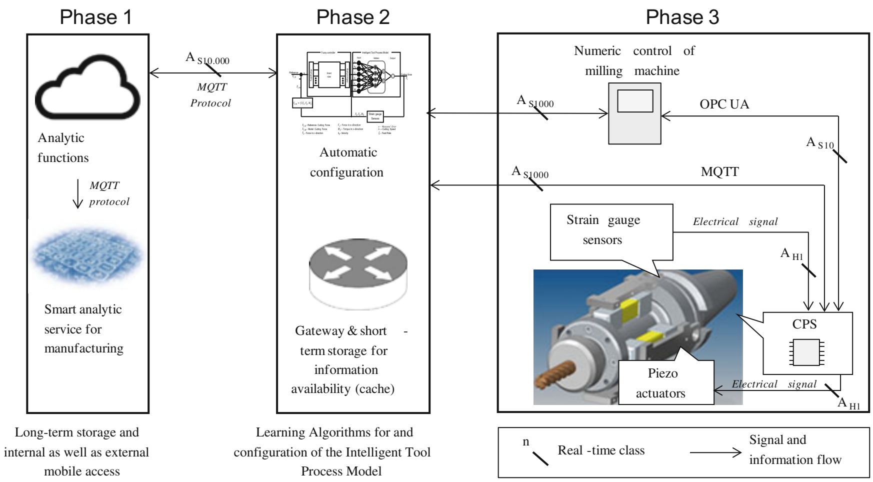


Fig. 6 CPS for milling machines

Therefore, an Eclipse Gura Framework in Java was implemented and a dumping database was introduced. In the event that the Internet connection is faulty or not available at all, the dumping database enables the caching of sensor data, which can be transferred to the cloud later on. The gateway communicates with the message broker. While the message broker subscribes to specific themes and issues, the gateway publishes sensor data. Therefore, an MQTT protocol approach was used due to its widespread importance in the "Internet of Things" space as well as the quicker response times in comparison to other concepts (such as https connections). The message broker is hosted in a cloud in the Internet. Thus, sensor data can be pushed to the cloud and stored in the database. These data can be visualized on mobile devices such tablets and smart phones. Therefore, mobile devices need to subscribe to the message broker in the cloud via any common Internet protocol. The control loop is closed by connecting to the smart device via OPC UA.

4.2.1 Phase I: Company Management Level—Smart Analytics

The infrastructure on the company management level enables the connection of smart devices and intelligent tools to company-wide databases or cloud services, respectively. It is possible to push these kinds of data which are rather of long-term importance and do not necessarily need a quick response time. The parent control loop favors non-real-time applications such as abrasion and deterioration of specific components, e.g. intelligent tools. Due to the rather slow, unreliable and somewhat delayed connection, the parent control loop predominately qualifies for soft real-time applications.

As already stated in the application of the intelligent chuck, several services can be provided to support the long-term analysis and functions. Predictive maintenance can be seen as an example of those services. Due to the data collection, an analysis can be conducted to estimate the wear behavior of the tools according to the parameters set by the machine control. Therefore, tool breakage can be avoided and predictive maintenance can be performed in time. Numerous services are conceivable in this context.

4.2.2 Phase II: Production Management Level—Automatic Configuration

Based on the infrastructure, it is possible to implement a controlling and management system. Therefore, this paper outlines a general controller strategy that can be tailored to specific use cases, such as the intelligent tool (phase III). An application for CPS-based modules is reasonable due to the wide diversity of the CPS modules, their increasing use in and the subsequently high complexity of modern production systems. The controller strategy is applicable for each phase of consideration (section on solutions in area of machining), which leads to flexibility and adaptability of the model. The approach is separated into a fuzzy controller, a neuronal network as a system, and an exemplary sensor for measuring mechanical forces.

The *controller* is represented by a set of fuzzy rules. The fuzzy logic has been applied in various fields, starting with control theory and extending up to artificial intelligence. While classic approaches are limited to Boolean values, the fuzzy method reasons based on partial and uncertain knowledge. The advantage lies in the possibility to describe uncertainty, whereas the value can range between completely true and completely false [14].

The *neural network* describes the corresponding system, such as the specific type of milling tool. It incorporates nodes that are connected and grouped together. A circular node displays an artificial neuron and an arrow from one to another represents the connection of an output to the input of another. The neurons are grouped in three layers: the input, the hidden and the output layer. In the example, the input has six neurons for the measured forces, the torque, the cutting force, the rotational speed and the feed rate. Each neuron includes an activation function and a constant weight that manipulate the output accordingly. The activation function can be binary, sigmoidal, Gaussian or multiquadratic. For a universal approach, the process model uses sigmoidal activation functions which have been proven as extremely efficient approximators. The weights are calculated according to the use case, such as the specific tool of the milling machine.

In the case of a milling machine, the *sensor system* provides information about nearby forces of the tool that are ideally close to the genuine cutting force. The sensor output is derived from the reference in order to calculate the measured error. The measured error represents the input for the fuzzy controller. An overview of the system is seen in Fig. 7.

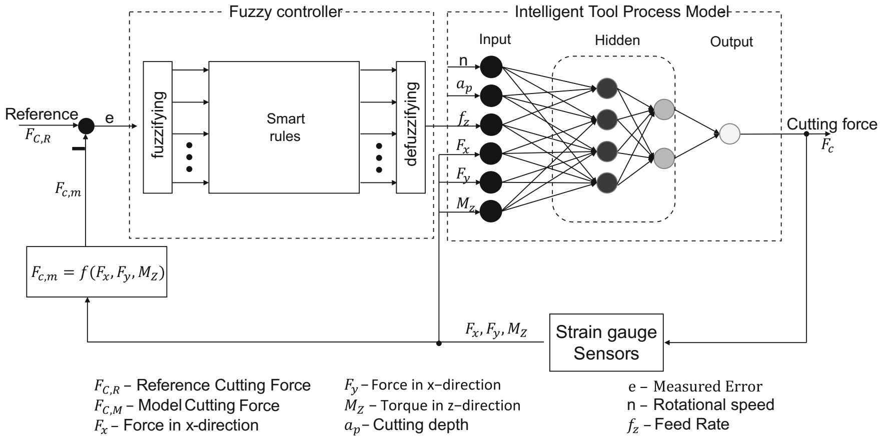


Fig. 7 Control theory

While the set of fuzzy rules remains merely constant, the neural network needs to be trained according to the specific type of milling tool. Therefore, various learning algorithms can be used in order to determine the weight of each neuron. The approach uses supervised learning, which aims to find an appropriate function that matches the training data. A cost function is thus applied in order to measure the gap between an optimal and an intermediate solution. In general it can be stated that the smaller the cost, the better the approximation to the training data and the better a solution can be expected. In the case of a supervised learning algorithm, the cost is represented by a mean squared error. The learning is iterative, meaning weights of the neurons are adapted and the error will be calculated between the output of the training data and the current solution in each step. The iteration stops as soon as the error is minimized across all sample points (gradient descent). As a result, the weights are adapted according to the training data.

The approach uses different training data sets according to the mechanical tool of the milling machine. Each tool has a digital representation in the form of the process model. The process model thus incorporates weights and activation functions of the thirteen neurons that are interchangeable in the controller in case of a tool change. Both the learning by the process model and the mounting according to the tool correspond to soft real-time requirements. Considering the IT infrastructure, the gateway pushes the process model onto the appropriate smart device. Therefore, the intelligent tool can be configured automatically using the appropriate trained neural network. In the example, the smart device inhabits the sensor technology as well as the logic to calculate output values and rules. These rules are again pushed through the gateway, which is in turn connected to the milling machine by an OPC UA protocol.

4.2.3 Phase III: Shop Floor—Development of the Intelligent Tool

Considering the evaluated potentials of the CPS in the machining process, it can be stated that the majority of the identified potentials depend on the knowledge of the applicable stress in the running process. To ensure high quality of the collected data, the measurement of the forces and torques has to be performed close to the force application point. Furthermore, the parameters of the machining process have to be provided by the machine control and matched to the measured forces and torques.

As already stated, the optimization of the productivity using an optimized stock removal rate under the consideration of actual process parameters is identified as a potential of the application of CPS. Considering the minimization of the safety factor, the threshold value of the applicable forces serves as the optimization parameter. In line with the increasing cutting speed, the excitation frequency approaches the Eigen frequency and a chattering of the machine may occur [3]. Consequently, this phenomenon must be taken into account during the development process. Therefore, two different concepts of CPS are involved in the development process to incorporate the compensation for the chattering and the optimization of the stock removal rate. These two concepts will be presented in detail in the following section.

Chattering compensation:
An active system has been designed which features data collection via sensors and a processing unit, including closed-loop control that activates the actuators used to compensate for the chattering [29]. An overview of the developed subsystems is shown in Fig. 8.

The acceleration sensors used detect the deviation of the tool center point via the amplitude of the tool tip. The self-induced oscillations that occur are transferred to the frequency spectrum of the acceleration signals, using a Fast Fourier transformation. Analyzed by the process unit, which is based directly on the tool, the

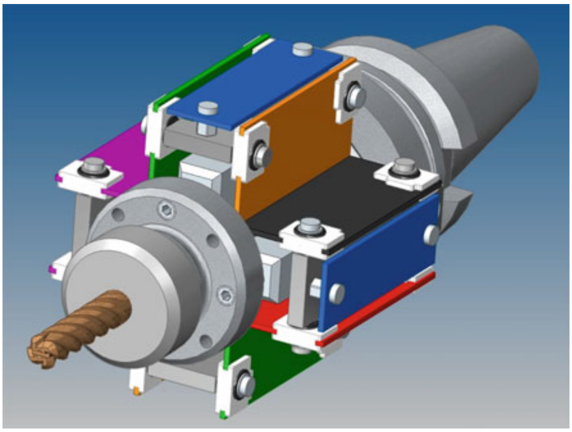


Fig. 8 Developed tool equipped with sensors for chattering detection

signals can be evaluated and processed to the actuators. Four piezoelectric actuators are used to compensate for the chattering, according to the tool center point. In the realized concept, the actuators initiate a 180° phase-shifted oscillation, whose frequency is similar to the excitation frequency of the chattering. Using this concept, the amplitude of the regenerative chattering can be dampened or even completely eliminated. Since the system will be equipped with a real-time operation system, the compensation can be used efficiently and the surface quality of the part is improved. A communication port is implemented to enable the data analysis, in order to acquire long-term dependencies and trends. Referring to the classification of the concept as an intelligent object, the system incorporates decision-making capabilities without an external control and the intelligence situated directly on the object.

Optimization of stock removal rate:
The concept for optimizing the productivity is realized via an adaptive feed control. A closed loop with the machine control needs to be enabled to ensure the impact on the target values such as stock removal rate, feed rate, etc. in dependency of the current application forces. The main distinction from the first concept is therefore the location of the actuator, which is situated in the machine control in the communication network and not on the tool itself.

Similar to the first concept, the applicable forces are measured via strain gauges close to the application point. The side loads, the axial forces and the torsional moment that occur are recorded directly on the tool holding fixture. By analyzing these variables, the load on the tool edge can be derived and the appropriate feed rate can be determined accordingly. The processing of the measurement signals and subsequent control is performed by a microcontroller circuit, which is integrated in the tool holder. The target feed rate is calculated on the circuit board itself. To ensure that the feed rate is chosen just below the maximum permissible tool load, the respective tool characteristics have to be taken into account in line with the calculation.

Since the placement of the feed rate is situated in the machine control, the calculated values for the adaption need to be transmitted in order to realize a closed-loop control. Due to the high noise immunity and low delay time, the transmission is conducted via Bluetooth. The relevant parameters will be adapted directly in the machine control. It can be stated that the intelligence is situated on the microprocessor integrated in the circuit on the tool itself. Referring to the classification of the concept as an intelligent object, the system incorporates decision making capabilities, including an external impact to interact directly with the process.

The consolidation of the two concepts in one tool would encompass the tradeoff between productivity and flexibility combined with high-quality requirements via the integration of electronic and information technology in a mechanic tool.

The following formula shows the real-time requirements for the pre-quality analyst function.

Formula: Source to Sink: $A_{H1} \rightarrow A_{S10} \rightarrow A_{S1000} \rightarrow A_{S10.000}$: A_{Sum} 11.011 s (6)

From this, it can be concluded that the necessary information is provided to the pre-quality analyst function in approximately 11 s.

5 Evaluation and Classification

First of all, the validation process of the developed CPS is demonstrated at the example of the closed-loop control. Furthermore, an analysis of the applied examples is conducted according to the specified characteristics and classification. The classification model is then applied afterwards.

5.1 Validation of the Developed CPS

The experimental validation has been taken place iteratively embedded in the development process of the devices, following a systems engineering approach. Exemplary, the experimental validation of the intelligent tool, in specific, the closed-loop control of the use case: "Optimization of stock control" is described in the following section to give an impression about the phases of the development process:

Phase 1: Development of CPS electronics and software
Phase 2: Configuration and test of the Neural Network and Fuzzy Controller
Phase 3: Experimentation in a genuine test environment of a milling machine

Phase 1: *Development of CPS electronics and software:*
The development of the CPS electronics and software was an iterative process. Various hardware prototypes were produced, altered and adjusted according for its application in a milling machine. As of today, the hardware is ready for the integration into a state-of-the art milling machine.

Phase 2: *Configuration and test of the Neural Network and Fuzzy Controller:*
The configuration and test of the neuronal network and fuzzy controller for the CPS needs experimental training data. Therefore, an experimental set-up of a machine tool has been conducted with two different tool tips and numerous variations of rotational speeds and feed rates cutting only one using three clampings. The results of the experimental tests are split in several sets to make sure that the data is significant and not co-dependent. The sets are used to either support the development and dimensioning of the closed-loop control or to validate the developed solutions. Regarding the concrete example of the closed-loop control, a two- step validation has been conducted in line with the development—a simulation-based validation of the neural network and of the fuzzy-control.

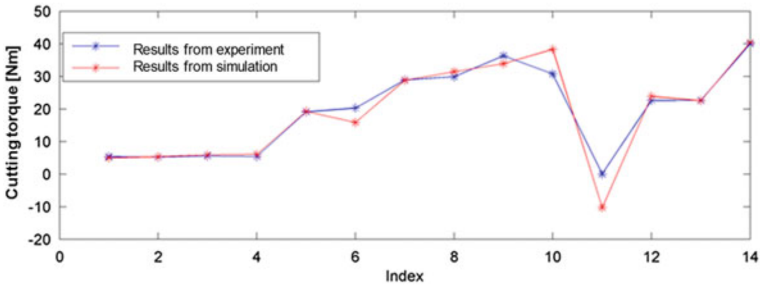


Fig. 9 Simulation results versus experimental results—Cutting torque

A configuration and test of the neural network have been executed via a software tool. The experimental data is therefore divided stochastically in three parts: The training, the testing and the validation. In the training phase, experiments will be conducted to define the number of layers and neurons. Via the variation of the simulation parameters as the training algorithms and the transfer functions, a comparison of the results with the experimental data is conducted. The best results are achieved with a resilient backpropagation algorithm and a logarithmic sigmoidal transfer functions. The maximum error rate of the torque is 2.2 % comparing the experimental and simulation results as shown in the following Fig. 9.

According to the former results, the configuration and test of the fuzzy control has been executed via a simulation as well. Using several variations of the scaling and the amplification factor, the simulation results of the feed rate and cutting force are examined and compared to the experimental results. It is shown that only a slight difference between a fuzzy PI-control and a fuzzy PID-control. Based on these examinations, the PI-control is chosen for this application according to the lower level of calculation effort.

Phase 3: *Experimentation in a genuine test environment of a milling machine:* While most of the work of Phase 1 and 2 is finished, Phase 3 has started and is still on-going. The CPS hardware and software will be used as a platform to implement the neuronal network and fuzzy controller. The simulation-based findings for the configuration of the neuronal network and the fuzzy controller will be integrated in the CPS hardware. The overall validation of the CPS system (closed control-loop) will finish with an integration test and will show the prospects and eventual shortcomings of the approach.

5.2 Evaluation of the Defined Requirements

An embedded system must determine the information flow and provides an essential function to the process.

A significant impact of the software, implemented on the embedded system, on the provided functions is the first characteristic. The process optimization occurs based on the supplied functionalities on a short-term and long-term basis. A shortening of the clamping or milling times as well as a possibility to ensure a more suitable configuration of the devices, based on historical data, can be seen as optimization potentials. Consequently, the functionalities of the software have a significant impact on the process.

Either horizontal or vertical communication needs to be implemented with the CPS.

The second criterion concerns the communication with the shop floor or production-based software systems. Since the two devices communicate either with a cloud or with a manufacturing execution system, the requirement is fulfilled.

Decentralized control of sensors and actuators must be implemented.

Analyzing the integration of sensors and actuators directly on the CPS, both applications comply with this requirement. The clamping device communicates the clamping forces via sensors directly to the PLC, to ensure a control loop. The milling tool combines actuators as well as sensors directly on the tool holding fixture.

Self-description as the knowledge about its own status, possible reference functionality and about the internal data must be incorporated in the CPS.

The last requirement of the CPS characteristics concerns the self-description capabilities of the CPS. Since the developed solutions are equipped with the necessary electronic and software to analyze the applied forces and resolve appropriate actions, the CPS are capable of describing their own status and possible functionalities. Furthermore, a communication concept is implemented in order to share this information with the respective software systems.

Adequate communication technologies must be chosen according to the real-time requirements of the process.

Communication technologies must be chosen in relation to their real-time requirements. According to this, the intelligent chuck contains three hard real-time requirements. The deployed communication technologies are provided on a wired basis using a fieldbus. The soft real-time requirements are realized via the OPC UA protocol or a web service, which provide more flexibility and functionalities. A similar scenario applies to the milling tool. One hard real-time requirement is required by the process, an electrical signal is used and a real-time operating system is implemented on the corresponding micro-controller.

The requirements of the characteristics and real-time are thus validated. The analysis and application regarding the classification model is presented in the following subsection.

5.3 Classification of the Developed Solutions

To display the differences between the CPS, the developed solutions are represented in the classification model.

The intelligent chuck incorporates decision-making abilities which are located in the network and not on the chuck itself. The vertical integration extends up to the company management level, since the impact is situated in the field of production planning and control.

The intelligent tool includes two concepts which are classified separately. The adaptive feed control has an impact on the company management level because it is used for analysis according to tool wear and breakage. The intelligence is situated in the network, as the applicable parameters can only be controlled by the numerical control. Active problem notification is the appropriate level of intelligence, since the operator has to implement a static process model in order to build up the configuration of the tool. The chattering compensation communicates only on the shop floor level, since hard real-time requirements arise in this system. The decision-making capabilities are directly situated on the object.

The visualization of the differences and the classification are shown in Fig. 10.

It can be stated that the CPS are rather different with respect to their application and their production environment. It is not always necessary to maximize the functionalities in each dimension in order to match the requirements. Nevertheless, all three dimensions of the classification model are applicable to the presented solutions. Therefore the validation of the stated requirements is conducted completely.

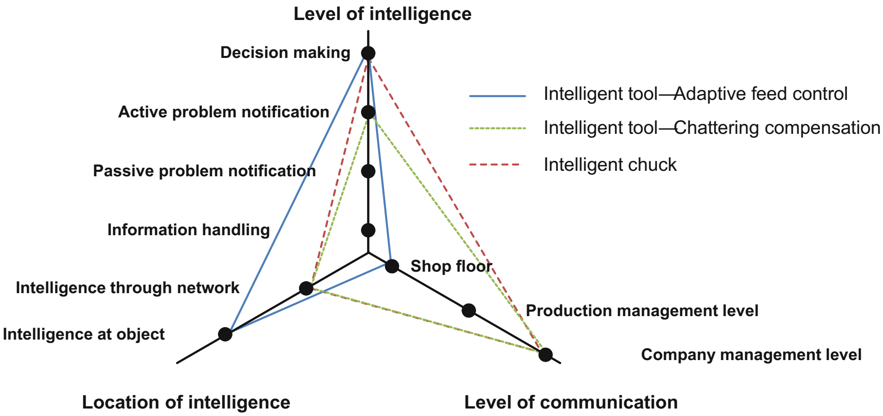


Fig. 10 Applied classification model

6 Summary

The growing importance of IoT in manufacturing raises questions with respect to its foundation, principles and application. This paper describes current challenges concerning the flexibility and productivity of turning, milling and drilling processes. The state of the art shows the potential of cyber-physical systems (CPS) in industry, but outlines the absence of technical as well as business concepts. A classification of objects on the basis of its intelligence forms the foundation for evaluating the developed solutions. As one of the major factors, the classification includes a definition of real-time which leads to the paper's approach to an information distribution concept for the manufacturing industry. The core of the solution includes three phases that are oriented on the automation pyramid. Each phase is outlined within the two use cases of an intelligent chuck for turning machines and an intelligent tool for milling machines. The intelligent chuck has the ability to adapt the clamping force on the basis of sensor data from the turning machine. The intelligent milling tool uses an IoT infrastructure to adapt and configure process models for tools as part of a controlling strategy to adapt the feed rate according to the force data of the specific milling tool. Furthermore, long-term storage is incorporated in a cloud service encompassing the ability to derive and display information about tool wear. Finally, the use cases are evaluated and classified based on the previously explained metric for intelligence.

References

1. Abele E, Reinhart G (2011) Zukunft der produktion: Herausforderungen, Forschungsfelder, Chancen. Carl Hanser, München
2. Bauernhansl T, ten Hompel M, Vogel-Heuser B (eds) (2014) Industrie 4.0 in Produktion, Automatisierung und Logistik. Springer Fachmedien Wiesbaden, Wiesbaden
3. Baur M (2015) Aktives Dämpfungssystem zur Ratterunterdrückung an spanenden Werkzeugmaschinen. IWB-Forschungsberichte, vol 290. Utz, Munich
4. Brecher C (2015) Advances in production technology. Lecture notes in production engineering. Springer, Cham
5. Broy M (2010) Cyber-physical systems: innovation durch software-intensive eingebettete systeme, vol 0. Springer-Verlag Berlin, Heidelberg
6. CyProS—Cyber-Physische Produktionssysteme (2013) CyProS—Cyber-Physische Produktionssysteme. http://www.projekt-cypros.de/. Accessed 11 Oct 2015
7. Geisberger E, Broy M (2012) AgendaCPS: Integrierte Forschungsagenda Cyber-Physical Systems, vol 1. Springer, Berlin
8. Kiritsis D (2011) Closed-loop PLM for intelligent products in the era of the Internet of things. Comput Aided Des 43(5):479–501. doi:10.1016/j.cad.2010.03.002
9 Klocke F, Wirtz G, Kratz S, Auerbach T, Veselovac D (2010) Modellbasierte und adaptive Regelung von Fertigungsprozessen. werkstatttechnik online
10. Lee EA (2008) Cyber physical systems: design challenges. In: Proceedings. 2008 11th IEEE international symposium on object oriented real-time distributed computing (ISORC). IEEE Computer Society, Los Alamitos

11. Lee J, Bagheri B, Kao H (2015) A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufact Lett 3:18–23
12. McFarlane D, Sarma S, Chirn JL, Wong C, Ashton K (2003) Auto ID systems and intelligent manufacturing control. Eng Appl Artif Intell 16(4):365–376
13. Meyer GG, Främling K, Holmström J (2009) Intelligent products: a survey. Comput Ind 60 (3):137–148. doi:10.1016/j.compind.2008.12.005
14. Novák V, Perfilieva I, Močkoř J (1999) Mathematical principles of fuzzy logic. The Kluwer International Series in Engineering and Computer Science, Secs 517. Kluwer Academic, Boston
15. Reinhart G, Wahlster W, Scholz-Reiter B, Thoben IK, Engelhardt P (2013) Cyber-Physische Produktionssysteme: Produktivitäts- und Flexibilitätssteigerung durch die Vernetzung intelligenter Systeme in der Fabrik. wt Werkstattstechnik online (2):84–89
16. Reinhart G, Wittenstein M, Scholz-Reiter B et al. (eds) (2015) Intelligente vernetzung in der fabrik: industrie 4.0 umsetzungsbeispiele für die Praxis. Fraunhofer Verlag, Stuttgart
17. Dorf RC, Bishop RH (1998) Modern control systems, 8th edition. Addison-Wesley
18. Sauer O (2014) Developments and trends in Shopfloor-related ICT Systems, International Automation Congress 2014
19. Sauer O, Jasperneite J (2011) Adaptive information technology in manufacturing, Conference: 44th CIRP Conference on Manufacturing
20. Schlechtendahl J, Keinert M, Kretschmer F, Lechler A, Verl A (2015) Making existing production systems Industry 4.0-ready, vol 9. Springer, Berlin
21. Schmid M, Berger S, Reinhart G (2014) Prioritisation of influence variables for a flexible and efficient production. In: Zaeh MF (ed) Enabling Manufacturing Competitiveness and Economic Sustainability. Springer International Publishing, Cham, pp 359–363
22. Schuh G (2006) Produktionsplanung und -steuerung: Grundlagen Gestaltung und Konzepte, 3., völlig neu bearbeitete Auflage. SpringerLink: Bücher. Springer, Berlin
23. Suh SC (2014) Applied cyber-physical systems. Springer, New York
24. Tarraf DC (2013) Control of cyber-physical systems. In: Workshop held at Johns Hopkins University, March 2013. Lecture notes in control and information sciences, vol 449. Springer, Cham
25. VDI 3633 Simulation von Logistik-, Materialflussund Produktionssystemen 03.100.10 (3633)
26. Ventä O (2007) Intelligent products and systems. Technology Theme: Final Report, Espoo
27. Venkata Rao R, Kalyankar VD (2013) Parameter optimization of modern machining processes using teaching–learning-based optimization algorithm. Eng Appl Artif
28. Vogel-Heuser B, Roesch S (2013) Eine Anwendung von CPS im Bereich der Wartung in der Produktautomatisierung
29. Waibel M (2012) Aktive Zusatzsysteme zur Schwingungsreduktion an Werkzeugmaschinen. Forschungsberichte/IWB, Bd., vol 273
30. Wang L, Törngren M, Onori M (2015) Current status and advancement of cyber-physical systems in manufacturing. J Manufact Syst 37:517–527
31. Wellenreuther G, Zastrow D (2005) Automatisieren mit SPS—Theorie und Praxis. Vieweg Verlag, Wiesbaden
32. Wong CY, McFarlane D, Ahmad Zaharudin A, Agarwal V (2002) The intelligent product driven supply chain. In: SMC2002: IEEE international conference on systems, man and cybernetics, p 6
33. Zbib N, Raileanu S, Sallez Y, Berger T, Trentesaux D (2008) From passive products to intelligent products: the augmentation module concept
34. Zöbel D (2008) Echtszeitsysteme: Grundlagen der Planung. eXamen.press. Springer, Berlin

Author Biographies

Christoph Berger, M.Sc. has been a research assistant at Fraunhofer RMV since 2014. His research interests lie in event-driven production control for cyber-physical production systems. Christoph Berger previously studied energy and automation engineering in Augsburg.

Juliane Nägele, M.Sc. has worked as a research assistant at Fraunhofer RMV in Augsburg since 2014. Her research interest lies in the evaluation of the impact of information and communication technology in cyber-physical production systems. She studied industrial engineering and management with a focus on mechanical engineering in Ilmenau, Berlin and Lyon.

Benny Drescher, M.Sc. has been a research assistant at Fraunhofer RMV/Institute for Machine Tools and Industrial Management in Munich and Augsburg since 2012. His research interests lie in the design of mechatronic machines, digital engineering tools and smart sensor systems. He has worked for major manufacturers in the mechanical engineering and automotive industry in Germany and Japan. Benny Drescher previously studied electrical engineering and information technology in Chemnitz, Munich and Cambridge.

Prof. Dr.-Ing. Gunther Reinhart is a full professor of Industrial Management and Assembly Technology and the Director of the iwb (Institute for Machine Tools and Industrial Management) at Technische Universität München. After studying mechanical engineering with an emphasis on design and development, he was a research assistant at iwb from 1982 to 1988 with Professor Dr. Joachim Milberg. During the last two years there, he was in charge of the assembly automation department. After receiving his Ph.D. from Technische Universität München, he started his industrial career with the BMW Group, initially as head of the handling and welding engineering department and subsequently as director of the body paint shop. In 1993, he returned to the university to become a professor and the Director of the iwb. He is also Chairman of the Bavarian Cluster for Mechatronics and Automation e.V., and since January 1, 2009, has been head of the Fraunhofer IWU Research Department for Resource-Efficient Converting Machines (RMV).

Going Smart—CPPS for Digital Production

Sven Goetz, Gunnar Keitzel and Fritz Klocke

1 Introduction

Global acting enterprises in the production area have to ensure all-time flexibility toward changing framework and boundary conditions. Ideally, the production plants are linked using information and communication technologies for data and information exchange. Therefore it is necessary that the production planning as well as the production and the assembly are capable of providing relevant data in real time ensuring short reaction times.

These data require a linkage between classical machinery and plant engineering with information and telecommunication technologies. These links will then provide companies with new opportunities for honing their competitive edge. The Industry 4.0 Initiative encompasses not only new forms of smart production and automation technology which will have a profound effect on the value-added networks, but also incorporates smart modelling along with production, engineering and production environment design which takes account of demographic changes and achieve new forms of work organization. It focusses on new technologies, models and systems that help the user to increase collaborative productivity in the interaction between humans and technology, provided the according framework conditions are given. However, in production industries, the mood is one of anticipation and certainty in equal measure. The future of industrial production is therefore defined by the answers to a series of questions:

- How can detailed models use semantic web technologies to interact efficiently with production in an intercompany or interdisciplinary environment?
- How can product and production complexity be kept under control in global supply chains?

S. Goetz (✉) · G. Keitzel · F. Klocke
WZL der RWTH Aachen, Steinbachstr. 19, 52072 Aachen, Germany
e-mail: S.Goetz@wzl.rwth-aachen.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_15

- What new services will emerge?
- How can cognitive knowledge be transferred to the digital world of production?
- Can Cyber-Physical Production Systems (CPPS) play a part in increasing competitiveness?
- What notable features will emerge from the global networking of customers, suppliers and production facilities?
- How can issues relating to data security and the protection of know-how be resolved to the satisfaction of all parties?

These questions were regarded in the past by the international community, for example the "International Conference on Cyber-Physical Systems" (ICCPS 2015) in Seattle/USA or at the "Aachener Werkzeugmaschinen-Kolloquium 2014" in Aachen/Germany by a magnitude of industrial and scientific representatives as well as politicians. Fertile discussions between all actors provided for some tentative answers. But they also revealed the tremendous amount of work that is yet to be done regarding the efforts to implement a competitive digital production. Consequently, the potential which could possibly be tapped via digital production networking will be explored in the following along with an analysis of the deficits currently affecting hardware, software and the personnel involved and what actions are needed in order to support Industry 4.0 in achieving a breakthrough within the framework of international collaboration.

2 Technology Knowledge for Digital Production

Cyber-Physical Production Systems (CPPS) are production systems that are composed of physical components (i.e. machine tools or provision systems) that are connected to each other via cyber systems that are able to comprise data relating to process and machine status as well as to material and product characteristics recorded by the integrated sensors. These systems are capable of summarizing and processing data to provide supporting information to the production and can interact with manufacturing equipment, products and people involved in the manufacturing process and with the digital world. This section focusses on the manufacturing industry and its possibilities to create and make use of CPPS, which form an important element in the Industry 4.0 campaign to explore new avenues in the effort to digitalize and network modern production. The Industry 4.0 campaign is based on the concept that the same information is used to plan, control and regulate the value-added process and is consistently available at all times to each operative level. In order to obtain this information, information gathering is a major subject. Section 2.1 will therefore discuss this aspect and show means of using sensors to gather, process and analyze production data. Automation engineering and stimuli from the field of information and communications technology are the driving forces behind developments. The section is hence divided into concepts of sensor integration and fusion on the one hand, and their applications in production environment on the

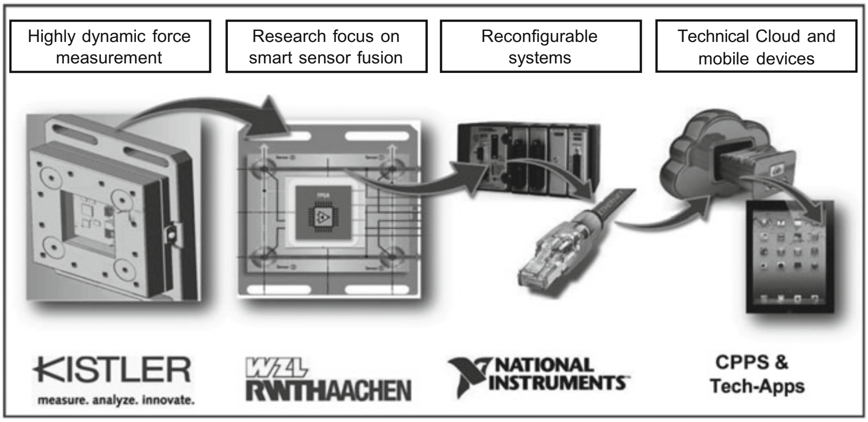


Fig. 1 From simple sensors to mobile use of information [7]

other hand. It shall help engineers be empowered to think holistically in the digital world of Industry 4.0 and to develop IT-based products and services. Section 2.2 will then discuss the use of this sensor information to feed Cyber-Physical Production Systems with data and information that allow the CPPS to operate. On the one hand, this implies the building and using of models to easily and cost-efficiently obtain and calculate the given information to create knowledge or to help humans in the decision making process. Since this is only viable if the created information can be brought to a human in a sensible way, the section also discusses Human-Machine-Interfaces and other Tech-Apps.

Figure 1 shows the data processing via all of these steps. Smart sensor technology helps obtaining sensor data, which is then merged with other sensor data into a full set of holistic information. This information can be used in reconfigurable systems to be fed into mobile and cloud devices and services. According to IEEE 1451, a smart sensor is the integration of an analog or digital sensor or actuator element, a processing unit, and a communication interface.[1]

2.1 Sensors—Perception Organs of CPPS

The Industry 4.0 campaign is about realizing the Smart Factory. The backbones of a factory are its production units, which are machine tools for many factories. The Smart Machine and possible ways of creating it are therefore the focus of this section. A Smart Machine is a CPPS, as will be shown in Sect. 2.1.1. Its smartness derives from its ability to sense its environment, draw conclusions from it and

---

[1]Lee [8].

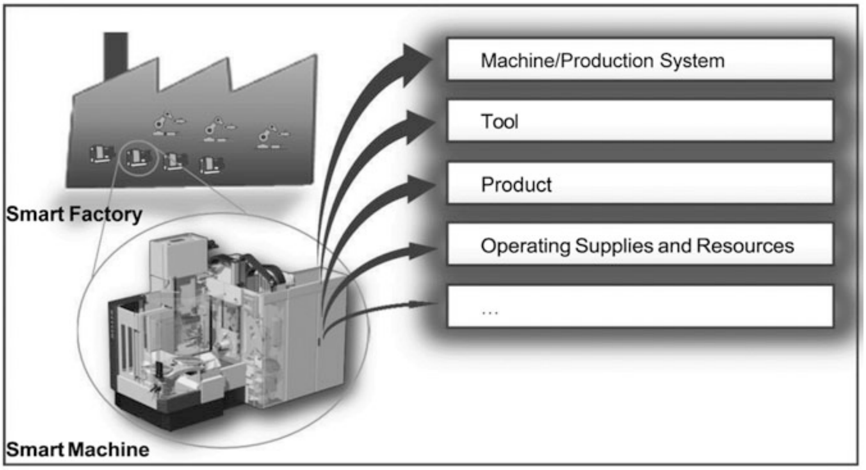


Fig. 2 Frame of reference in production [5]

communicate it to its surrounding world. In order to be able to sense its environment, the Smart Machine therefore needs perception organs, which are sensors and sensor systems.

Figure 2 shows the dimensions of a Smart Machine. The machine tool is only one aspect to a smart machine. The tool, product and especially all supporting units are also to be considered when designing a Smart Machine. All these are areas from which to extrapolate technological expertise:

- machines as smart units, which are networked with other machines and systems within the production environment,
- tools incorporating monitoring functions and information technology which enable them to recognize conditions, generate tool information and recommend process settings,
- smart products containing information about their own quality and part functionality,
- efficient, optimized use of consumables and resources in manufacturing processes and machine tools.

2.1.1 Concepts of Sensor Integration and Fusion

As stated above, CPPS are made of combining physical production with cyber communication. CPPS are independent systems that are able to autonomously gather and process data, communicate them to their environment and make decisions based upon this information. They can be described as a system of networked sensors based on independent fusion. The network relies on adaptive system information passed by the CPPS to the sensor system. The Cyber-Physical Sensor System takes account of the optimization goals and supplies the process information required for the development of suitable process models to the CPPS. The required information is drawn from the raw data before the process or individual part characteristics can be monitored independently.[2]

In order to do this, CPPS need algorithms that can be adapted highly dynamically. Not only horizontal information exchange is required for running CPPS, but also vertical connection to superior or inferior information levels, such as PLM systems on the planning level. Only by this connection both vertically and horizontally can the required level of adaptivity be achieved. Future process monitoring systems will be configured via production planning. The optimization criteria can either be filed via characteristic values[3] or they can be determined via modelling in CAx systems. To decrease calculation times, characteristic values may be used. This is especially gaining for SMEs which are commonly struggling for resources. Characteristic values have the advantage of reducing both the required computer performance as well as the volume of data involved. In extreme cases, characteristic values can be described simply via attributes as simple as “valid” or “not valid”. This concept goes along well with the concept of decentralization, where each (decentralized) CPS uses its own internal data processing to pass forward not the raw data of a measurement, but this very characteristic value. To raise efficiency, each entity is only provided with the information it needs. Referring to the Kanban philosophy, each CPPS will therefore “draw” the information it needs, as opposed to todays’ philosophy of providing or “pushing” information. Otherwise, there is a danger that volumes of unused data will accumulate as they are filed to server systems but are never used. The customer of a manufacturer will require information only as to whether a certain part fulfills all relevant quality attributes. Nevertheless, attribute evaluation and documentation do not suffice on their own for a “smart” manufacturing process. In order to guarantee product liability and certification requirements, it is important to file quantified documentation of important process parameters within the company. To realize such a data consistent quality assurance system, a company needs to develop strategies to manage and archive measurement and quality data in a sensible and efficient way over many years. Meta data play a major role in this aspect, as they are vital for structuring the Big Data and making a later request for a certain piece of information as well as its interpretation a lot easier. The limiting factors are the transfer rate, actualization times of the higher levels and the latency of modern bus systems in machine tools. The connection of the Smart Machine to several, different control levels within the company increases the demand for real time data and information. This will also increase the demand for real-time processing capacities. Nowadays, data processing was mainly executed in centralized computers. The next generation of sensor

[2]Langmann [9].

[3]In this context, a characteristic value is the merging of many data to form it into a condensed piece of information.

systems will have to be capable of processing data autonomously. The information supplied can then be made available to actuator components or to a higher level system of targets for further processing. This is one major attribute of CPPS. But not all processes require a real-time surveillance, and not every ERP system requires real-time data. Non-time-critical applications may therefore still be provided with database systems via standard network protocols. A CPPS sensor without a host could process even a direct stream of data to a decentralized data storage device. The CPPS will hence have to decide which data it processes autonomously, and when to forward information to a centralized database to be stored and distributed from there. In order to maintain a degree of flexibility as high as possible, it will therefore be essential to deploy more reconfigurable and scalable systems based on real time processors or FPGA technology. Since customer requirements vary significantly, most manufacturers of sensors and actors can only provide one hardware platform. Thus, implementing and using CPPS calls for the development and spread of a uniform, open and embedded sensor platform that lets new sensor systems be easily integrated.

Such an open embedded system would permit its users to develop and facilitate their own algorithms and add new functions to measuring devices and actuators. In the entertainment industry, this philosophy of providing an open and embedded platform to allow individual development of applications has already emerged, due to the customer demand of smart products. These smart devices can be adapted via “tech-apps” to meet the requirements of individual customers. Due to the openness of the platform, these apps may be programmed directly by the users who have a certain need, and then either be sold or licensed to user users. Future terminals are open, reconfigurable platforms which can be adapted by the customer and used to form communities. Through the development of Tech-Apps, these user communities generate new areas of applications and functions for the manufacturer.

But gathering and processing data is only one aspect of establishing an integrated and fused sensor system. The greatest challenge is how to manage the data and information that is generated in the CPPS. Meta data was already mentioned in this context above. Adding descriptive information to the measurement data helps finding what is needed, and furthermore also understanding the context in which a specific piece of data was generated. These aspects will be discussed in the following Sect. 2.1.2.

2.1.2 Applications in Production Environments

Standardized data formats provide an ideal data structure which permits data management systems to be developed rapidly and cost effectively using commercially available systems, without sacrificing the classical functionalities of a database. This development allows for the possibility to find data even from an outside access. Emerging technologies in the Big Data field, such as cloud systems, are supportive in that matter. Cloud Computing is the term used to describe the approach of providing dynamic IT infrastructures, calculation capacity, data storage, network capacities or services on demand. The advantage for companies is that resources which are required for only a short time do not come hand in hand with irreversible, costly expansion of the IT infrastructure. The data gathering and manipulation process was already shown in Fig. 1, where it shows the direct transfer of data to the cloud or to mobile devices.

So far, sensors that are available to the market and thus are used widespread the industry in real world modern production are only very limitedly usable for only applications. The following will therefore show the limits of todays' sensors and outline new sensor concepts. Without information as to the position of the work-piece and tool, any attempt to evaluate the signals emitted by a force sensor with a view to measuring process forces ends in misinterpretation. Hence, modern sensors need an holistic approach and a well-connected environment to be able to draw on all relevant information. In addition to this, individual sensors may measure process variables and send signals to assist in-process monitoring but extended signal processing is always required in order to permit model-based interpretation of the quality produced or of the condition of tools, machines and auxiliary devices. From these thoughts, two strategic directs can be derived in research in a drive to close the existing gaps, as shown in Fig. 3. One direction describes integrated sensor solutions being capable of supplying a higher volume of information. These will be sensors which already permit evaluation of the data which has been recorded and which, in conjunction with suitable models, already supply information instead of simply passing on signals to the next level—e.g. machine control, the operator or to the process planning level. The other direction is given by multi-sensor systems that measure several quantities in the same system. These sensors are actually sensor networks or embedded or integrated solutions which permit several measurands to be measured with the same sensor system. Sensor fusion describes the pursuit of both directions simultaneously. It describes combined smart signal processing by a multi-sensor system of several measured variables.[4] This is shown in Fig. 3.

Especially in machining operations, process forces are of high interest to the user, and their measurement is therefore essential for process analysis and process optimization. In addition to temperature models, force models are frequently applied approaches in order to optimize cutting operations and to achieve the required level of quality in complex products. The most common device to measure process forces is based on piezoelectric effects. Individually pre-tensioned quartz disks are connected to a sensor with high linearity. Additional information can be achieved by interconnecting several sensors to a force measuring platform. This does not only expand the range of the measurement, but also permits calculation of torque values. However, external effects, such as temperatures or gravitational effects will influence the otherwise linear behavior of the piezoelectric elements and make a correction necessary. E.g. the temperature of the elements causes a change in the correlation between force and the charge given off by the sensor. It is therefore essential to characterize this behavior and to calibrate it accordingly prior

[4]Klocke [5].

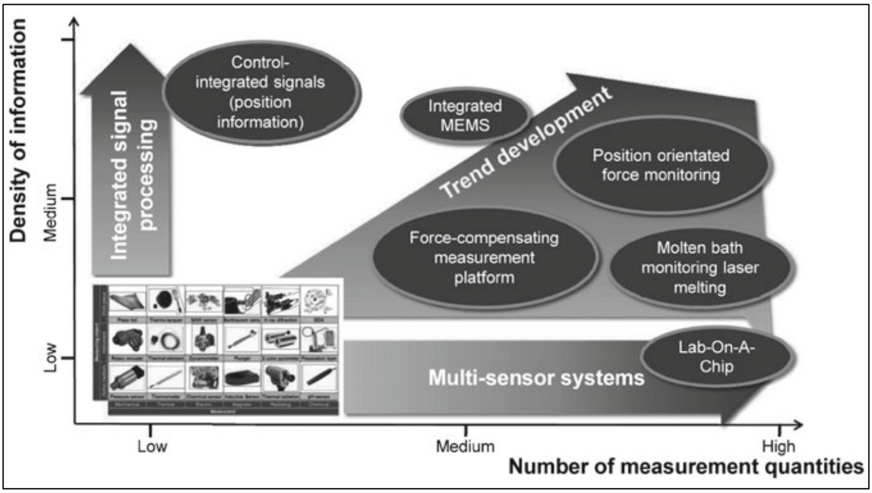


Fig. 3 Trends in sensor development [7]

to the manufacturing operation. But not only mechanical compensation by a balancing construction takes place. Integrating other sensor signals into the force measurement system can greatly enhance the systems' performance. For example, thermal effects can be corrected with a view to increasing measuring accuracy. If, in addition to the actual test signal, information relating to the operating temperature is available, a model-based correction can be carried out, thus further increasing the accuracy of the measurement. Compensation methods like this are used both to measure the cutting force and within the machine tool itself.

Temperature is a pivotal variable in the cutting zone and can impact directly on both tool wear and product quality. Process input variables such as the material to be machined and the cutting material or process parameters like cutting edge geometry, feed rates and cutting speeds exert considerable influence on the temperature in the machining zone. While at low material removal rates, temperatures are usually low due to low process forces, higher productivity may arise the problem of temperature induced surface failures, such as white layers or carburization. Depending on the optimization target, the ideal window within which the process should be performed may involve various cutting zone temperature ranges. In broaching processes, one possible approach to this dilemma is the addition of complex sensors to the process in order to control the temperature during the process to keep it below certain levels. The measurement setup equipped with the high-speed thermal imaging camera (SC7600) was supplemented by a 2-color pyrometer (2cp). By referencing measurement signals internally from two neighboring wavelength areas, the 2-color pyrometer provides a means of eliminating the influence exerted by the emission ratio and, thus serving as an optical, absolute temperature measuring device. The inclusion of the pyrometer signal allows the thermographic measurements to undergo a dynamic, temperature-dependent

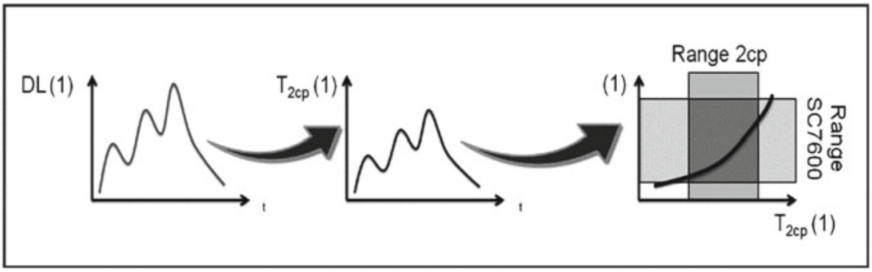


Fig. 4 Concept for the calibration of infrared camera data [7]

calibration. Figure 4 shows how a calibration function is generated via temporal synchronization for the work area overlapping the two systems.

One challenge when setting up this measurement system is the difference in the sampling rates of the measurement signals. While the 2-color pyrometer provides measurements at selected points, the thermography heat radiation is recorded over a relatively extensive image section. Furthermore, the total measurement range is far more extensive, beginning with ambient temperature. Therefore, integration times are longer and the sampling rates diminish at the same rate. Hence the time intervals between two measured points are in disparity. In order to identify a suitable calibration function, it is therefore essential to include the position of the pyrometer when process data are used.

The calibration curve in Fig. 5 shows a significant improvement of the results. This was achieved by adding information referencing the location of the 2-color pyrometer. These measurements can also be calibrated by merging the two measurement sets. The temperature distributions over the process are absolute, providing all of the data required for further process modelling operations.

This merge requires two measuring systems that both give temperature information. One signal gives an absolute temperature, while the other gives the temperature distribution. The systems differ in terms of their temporal and spatial resolution capacity. It is conceivable that there will be other production engineering

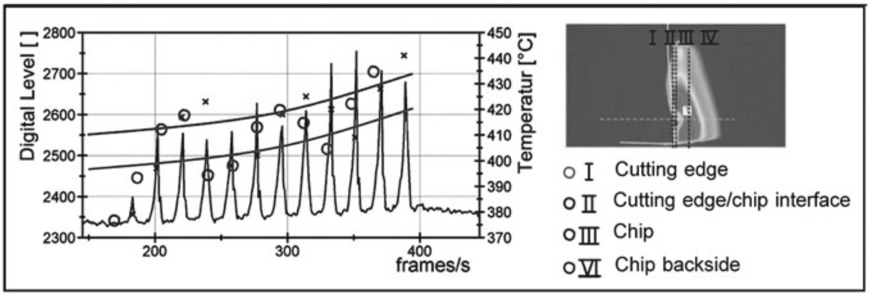


Fig. 5 Reducing calibration variance via location-referenced measurements [7]

applications which will require sensor fusion in order to create an innovative sensor with completely new capabilities by combining temporal or localized resolution capacity.

In production, the smallest unit in the manufacturing process is a machine tool or a work station. Data either available within the machine tool or generated by it can provide precise information about the manufacturing process and the condition of the machine tool as well as process status and conditions. This information represents an excellent database for the analysis of processes and machine components. Consequently, technological knowledge within a machine tool and its control unit can make a major contribution to digital production.[5]

Since sensor technology is very costly and many sensors are still designed to work in laboratory environment, new ways have to be found to gather data in a low-cost manner in an industrial environment. In smart machine tools, data relating to the process forces can be obtained from measurements of the current power consumption of the motors powering the drive axes. One of the most important research issues to be addressed is whether information, which could be used in order to optimize methods of monitoring tool wear and collision avoidance, can be obtained from the control signals. Ongoing research identifies which signal sources within the control unit and information lend themselves to supplying a cluster of machine-independent signals. Parameter models to compensate friction and acceleration effects will subsequently be implemented. The parameters typical of each machine will be generated autonomously in a routine operation. All insights gained and methods developed on this basis must be designed to be applicable to a range of machine systems, c.f. Fig. 6.

Regardless of where the sensor data comes from, the systems must be coupled to higher-level CAD/CAM systems and material databases to use them in technological surroundings. Communication between these systems and databases must run fully automatic in future. Using models instead of empirical data can help this task, as they tremendously decrease calculation times when compared to extracting information from a set of empirical data. One example is wear models such as those used in broaching operations. They can exploit considerable potential for increasing efficiency, converting it into productivity when expensive special-purpose tools are used. This is to be observed in the manufacture of safety-critical parts for aircraft turbines. Here, tools are frequently replaced long before their performance capacity has been reached. The reason is the lack of knowledge about the tool's exact state of wear. This leads to far higher tool costs than in the ideal case where each tool is run to its individual end of life. The development of a suitable wear model analytically formulated on the basis of empirical knowledge and correlating current tool temperature with process forces and state of wear can provide direct feedback from the process via in-process temperature measurement.[6] This leads to being able to run the tools up to their individual end of life, saving a considerable amount of money.

[5]Klocke [5].

[6]Rudolph [11].

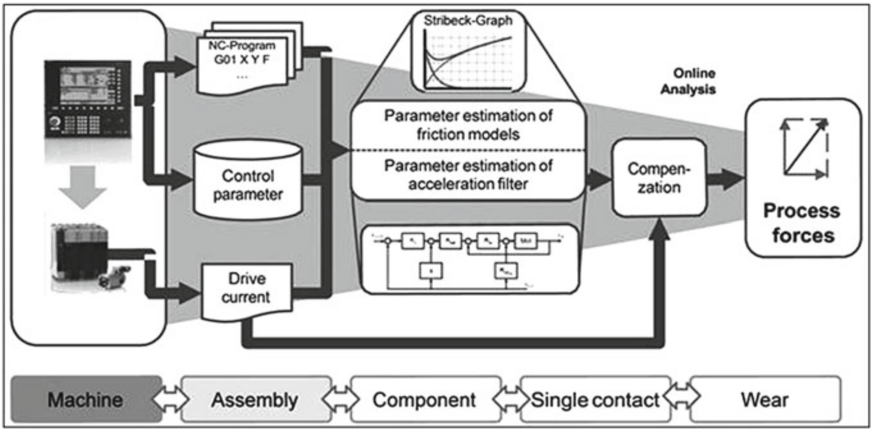


Fig. 6 Determining process forces from data within the control unit [11, 12]

When this wear model is extended to a control system, the user can define his or her own cost-function to control the process between the two extremes of high productivity and high tool costs.

Yet one more example can be drawn from complex hobbing operations. In an analogy experiment, the contact conditions can be abstracted and hence be analyzed in much greater detail. The load to which the tool is exposed is determined via sensors. This was done in a gear-shaping process where machining forces and the local temperatures were measured. Additionally, the chip formation was captured using a high-speed camera. The fusion of these three sensor data allowed for a deterministic approach to analyzing the contact conditions.

But practical experiments also bear some limits: They are time consuming, often costly due to sensor equipment, and in some cases the measured data underlies uncertainty and the process cannot be measured at the exact point where it would be necessary. Hence, numerical calculations have reached an accepted and required status modern production science.

For example, they are carried out in order to simulate the load to which the cutting edge is exposed. A process model[7] to predict tool life at the specific level of load involved has been developed on the basis of calculation and investigation. These results have been used to develop a process model based on a geometrical penetration calculation, which has achieved a considerable reduction in the calculation time required.[8] This process model builds the basis for a coherent and holistic process monitoring system.

Designing and manufacturing smart products calls for the implementation of workpiece-orientated monitoring. This is explained on the example of rotating,

[7]Klocke et al. [6].

[8]Herzhoff [3].

safety-critical areo-engine parts, such as turbine blisks. They are particularly demanding for to two reasons: For once, their environmental conditions (high temperature, high strain) require them to be made of hard-to-machine material. This creates very high demands to each production operation involved in their manufacturing process. Furthermore, turbines underlie the highest safety standards and are frequently observed by local and international authorities. This is done in order to rule out any failure of these safety critical parts when they are in operation. The current best practice is to screen part characteristics using Low-Cycle-Fatigue (LCF), High-Cycle-Fatigue (HCF) and Thermo-Mechanical-Fatigue (TMF) test methods. These tests are conducted after manufacture and involve the destruction of the components. The primary goal of in- process testing is to provide indicators as to the part characteristics from significant process information. This encompasses in-process evaluation of the finished surface as well as an assessment of peripheral rim damage. When faults or critical conditions are identified at this point, there are two options: Either conduct a subsequent machining operation or manufacture the part again from scratch.

This points out the need for a continuous data generation throughout the course of all manufacturing processes. Such procedure facilitates the assessment of conditioning the product over the course of time. Especially for safety-critical parts, but also for quality and customer satisfaction reasons, a consistent data acquisition and its correlation to a unique item creates higher value for the manufacturer as well as for the customer. Only when this has been achieved in the case of safety-critical engine components can the manufacturing process and the products be described as digitalized. In a digital production environment, the product becomes a smart product as a result of coupling manufacturing and product data to the part.[9] However, the data needs to be compressed for longer storage requirements. As more and more process information is gathered, this aspect becomes even more important.

Besides the actual process monitoring, each cutting process also needs consideration of supporting activities and all items that ensure this support. They can be classified as consumables. Consumables include water-mixed cooling lubricants, oils in cutting and forming operations, dielectrics in electrical discharge machining and electrolytes in electro—chemical machining operations conducted on metallic materials. They can be characterized in laboratories and tested either automatically or manually at both regular or irregular intervals. However, this characterization is usually done offline and not continuously, so that a direct feedback of its condition to the part quality is not possible. Yet there is ample evidence that the condition of those consumables has tremendous effect on the quality of the manufactured products and on the productivity of the manufacturing process. Status data are not available for direct processing in models. One reason can be seen in the difficult and costly gathering of this data. A solution is given by the use of model-based, miniaturized analysis systems, such as lab-on-a-chip systems. They permit consumables to be characterized in terms of age, chemical composition or

[9] Hardjosuwito [2].

contamination and are currently under development. As they are working online and parallel to the running process, they might be used in real production environment and allow for a direct link to the exerted influence of the consumable on the production system. As all sensors are integrated onto one platform with an internal data processing unit, they can be defined as CPPS.

2.2 CPPS—The Architecture for Smart Applications

The previous section discussed the usage of CPPS in a production environment and gave some scenarios in which and how to implement CPPS into manufacturing processes. It can be conducted that Smart Machine Tools and CPPS are connected by an interlinked and model based sensor system that creates and processes data autonomously on a decentralized basis. According to the definition of CPPS, those sensor systems can be called Cyber-Physical Sensor Systems (CPSS). This architecture is shown in Fig. 7.

The following Sections have a look at the architecture of this construction. Section 2.2.1 briefly outlines the usage of technology models that help reducing the processing demand of CPPS and CPSS, while Sect. 2.2.2 discusses how processed information from CPPS can be forwarded to its user, the human. This requires a smart HMI.

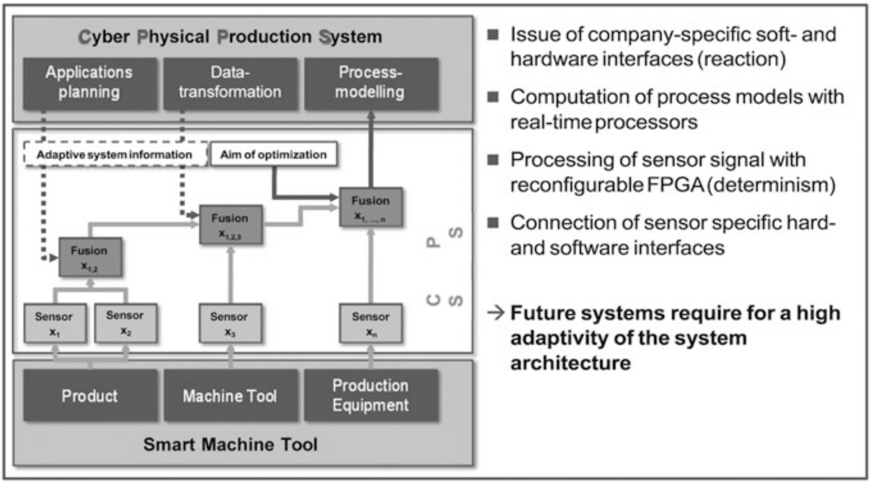


Fig. 7 Cyber-physical sensor system [7]

2.2.1 Technology Models—Knowledge Carriers of Production Entities

Increasing requirements nowadays challenge the manufacturers of metallic parts. The problem which occurs most often is the demand for reproducible manufacture of geometric forms. Form deviations can arise as a result of variations in the properties of the semi-finished product or of tool wear. Simulation plays an important role in this, but calculation capacities pose a bottleneck for high-performance simulations or real-time solutions. Therefore, a model-based approach might be able to significantly reduce calculation times by creating simple approaches to complex processes. Such models would also produce sets of process parameters for production processes. Normally, the process parameters are adjusted manually on the basis of production scenarios or of previous deviations revealed during the target and actual comparisons. The selection of new parameters then depends on the experience of the machine operator. This results in a protracted and expensive process which occurs at each phase in the process life cycle. Additionally, process requirements are becoming increasingly demanding due to the general trend towards miniaturization and reduced tolerances coupled with increasing material strength. This procedure can be shown in one example of a bending machine.

In an effort to reduce the reject rate and the tooling times, a model-based approach was selected for an adaptive control strategy. This firstly involves modelling the production process, for example a bending process. The bending process is initially analyzed by varying the process variables which exert the greatest influence on the process. This is achieved via corresponding simulations. The correlation between the significant variables and the geometrical deviation is determined and various self-optimizing control strategies are developed and tested. A special-purpose tool was developed in order to validate the simulation and to test the quality of the self-optimizing control strategy. This tool has an additional measuring device and can be used on standard test machines. When the self-optimizing control strategy was tested under real production conditions, the process parameters of interest in this case, the initial dimensions of the product to be bent were kept within the tolerances, thereby achieving a reject rate of 0 percent.[10]

This example shows the potential of models for reducing calculation time. It is achieved by expressing a complex problem with simple analytical equations that are easy to process. This process often requires the developer to firstly undergo a simulation to learn about the system behavior, and then to find simple equations that erase unimportant but costly aspects of the real world or that fit the real result sufficiently. Especially for phenomena that merge several different physical aspects (so called "Multi-Physics Systems") there is a large potential for reduction of effort. One example for a large reduction of effort by using a simplified equation is the cos (ϕ)-method, which has been applied extensively in industry for a number of years. It is used for complex and non-linear processes, such as electro-chemical machining

[10]Pitsch [10].

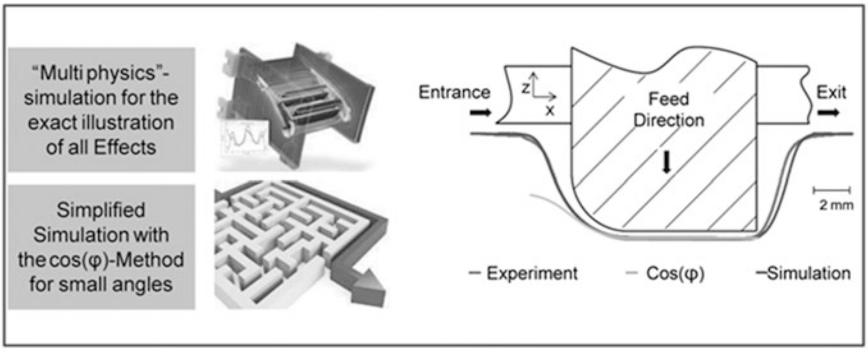


Fig. 8 Modelling concepts for reducing calculation times [5]

(ECM).[11] Depending on the specific issue in question, simulation approaches which have been optimized in terms of cost-effectiveness can be pursued individually for each specific application. Multiphysics simulations draw on all physical and chemical phenomena in modelling and are therefore capable of describing the relevant effects over the entire geometry in precise terms. Fluid dynamics, the electric field, the actual electro-chemical material removal and process-related modifications to the properties of the electrolytes can thereby be taken into account. These approaches, however, require considerable computer capacity and long calculation times.

The cos(ϕ)-method is used extensively in industry and offers a large potential, since it models the removal behavior with sufficient accuracy in the near region of the tool, c.f. Fig. 8. For small angle deviations ϕ from the feed direction, the machining behavior can be estimated by simply projecting Faraday's law in a significantly shorter calculation time.

Additionally, in the future it will be possible in the future to reduce calculation times for more detailed models using global IT resources, such as cloud computing. This permits large numbers of process data evaluations to be carried out at once. Distributed calculations in the cloud will be able to draw on computer resources, enabling them to optimize processes swiftly and individually for specific machining tasks. Prominent examples of such ideas are expressed in projects like "Seti@-Home" or "World Communication Grid", where even private individuals grant processing power to a globally integrated network. They can be seen as prototypes of industrial distribution systems that have yet to be developed. However, in the case of industry and company specific implementation of application-oriented products in the field of manufacturing technology, confidentiality and security aspects along with the regulation of rights and obligations of all parties involved and whose interests may vary, are essential. The example outlined reveals how it will be possible to file fundamental technological understanding on in-company and global IT structures. There is still a need for technology experts to research and

[11]Hinduja and Kunieda [4].

implement complex modelling approaches. However, it is vital to ensure that these models are available at the production planning stage and can be used at all levels in a modern production environment facility.

2.2.2 Tech Apps—The CPPS Human Machine Interface

Although many improvements in the field of CPPS support automation of production processes, the human in his or her role of a flexible and adaptive decision-maker will still play the major role in highly industrialized factories. Due to an increase in the complexity of many processes, workers on the shop floor often struggle to keep pace with the advancements. This implies two reasons for connecting the human to CPPS: The first reason can be seen in the flexible ability of humans to reach decisions. The higher the level of information the worker has, the higher the quality of his or her decisions. Transferring the information from diverse CPSS to the worker will enable him or her to reach to the best possible decisions. The second reason is the enhancement of the worker's skills. Augmented reality or assistance systems are examples for such a use of CPPS for the support of workers. In both cases, a simple and intuitive access to the relevant information as well as a removal of all unnecessary information is vital for the utility and acceptance of the system. Human-Machine-Interfaces (HMI) is therefore an important topic when discussing CPPS. The following will briefly discuss some aspects of developing HMI.

The term app (application software) is used to refer to application programs designed to provide useful functionalities on a system. Most apps are aimed at end users and are used for mobile devices such as tablet PCs or smart phones. The functionality of an app is limited and provides a defined service. Their operation is intuitive and no expert knowledge is required to install an app. One consequence of this is that consumers frequently use apps for only a limited period before switching to a different one. This trend is boosted by the intuitive operability of such apps. The availability of a wide range of apps in the web-store means that customers can decide spontaneously to install an additional functionality on their device. It is a logical progression to develop these tools for production environments too. We are still in the early stages of this process; there is enormous scope for collaborative research and implementation in an international context.

One advantage of technology apps for mobile devices is their large acceptance in the work force. Mobile devices are so ubiquitous in society (c.f Fig. 9) that virtually all employees are accustomed to using apps. This in itself provides enormous potential for improving communication in a digital production environment. Using those IT tools supports the distribution as well as the exchange of information in the production environment. The utility of social apps lies greatly with their intuitive use and them being designed for specific purposes. To make tech apps as powerful as social apps, these aspects need to be considered when designing them. Apps should have a slim user interface and be designed for specific technological purposes. However, to ensure that the worker is not overwhelmed with manifold apps,

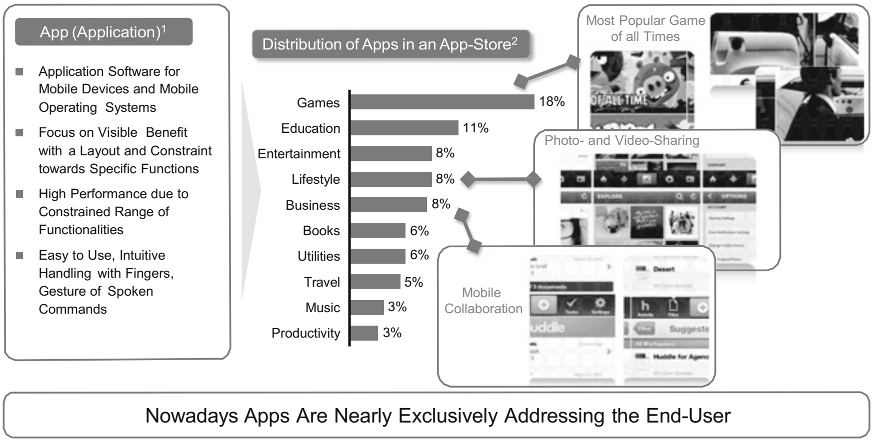


Fig. 9 Apps in mobile communications [10]

different technological aspects may be merged into the same app, as long as they can be called upon via the same user interface.

Aside from the advantages of tech apps mentioned above, they also have the potential to contribute to another goal: Their necessary connection to worldwide networks allows the exchange and acquisition of global technology knowledge, thus making it generally available. One of the requirements of tech-apps is that these applications are capable of providing the operator with clear feedback about the process. It is just as important that contextual support is available to the operator in cases where there is a high level of variance. It is anticipated that where there is a need for additional information, the tech-app will have access to IT infrastructure in the cloud and will use the process models available there. However, real time diagnosis of process status relies on a high-performance IT infrastructure. The interaction between humans and machines with support from tech-apps is presented in Fig. 10.

However, the requirements of tech apps exceed those for social apps tremendously. Not only do they have to meet the demands of the society (i.e. the workforce of a company), they also need to present an industry-specific solution which in turn also meets the requirements of the company concerned. If used in a network of suppliers and customers, the requirements are even larger, since external communication and safety issues have to be taken into account as well.

But when designed and implemented correctly, future operators faced with process analysis tasks or with machine tool operation will be able to access support based on technological models coupled with an understanding of the cause and effect relationships between tool, workpiece and machine tool. Additionally, the connection to the World Wide Web allows for even more new approaches to machine operation. For example, a remote machine communication would allow a worker to increase his radius of action within a factory, as he is always informed

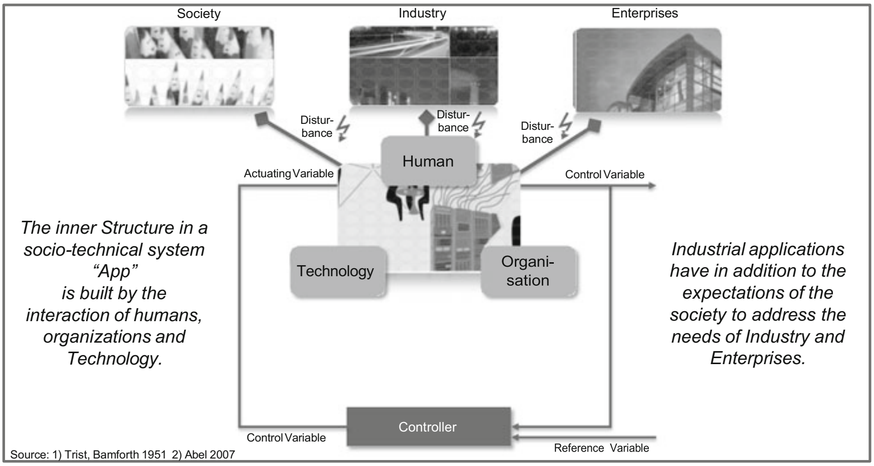


Fig. 10 Straightforward interaction between humans and machines via tech-apps [5]

about the condition of his machines or its processes. This could even go so far as to allow for external supervision by experts all around the world. Status messages can be sent to smart phones, certain machine functions can be controlled remotely via iPad and on-going machining operations can be optimized by an experienced operator who is not on-site.

However, there are still obstacles before making use of such systems. Many of the process models behind modern CPPS have to be further developed until they reach a greater technological maturity, and the machine tools as well as the running processes require higher process stability and reliability independently and without active control on the part of the operator. This does not apply for automatic process control, which contrarily would contribute to the process stability. A virtual, digital production must run alongside the actual production process to permit online comparison between the target and the actual status of the process.

Thorough, technological understanding is the most important requirement for digitized production. A process can be performed independently only when it is in an environment of thorough and extensive understanding of the process. Technological know-how must be implemented as tech-apps and in more complex simulation models for the cloud so that production systems can access and use this expertise.

One example for such an app is shown in Fig. 11. While there is a collection of models that calculate the cutting force in a process with geometrically defined cutting edge, complex processes cannot be reproduced using simple tools and machine operators certainly cannot be expected to work out which process boundary conditions prevail at a given time before using suitable models to draw conclusions as to the cutting force or the moment required without software support.

Tech apps, however, can solve this matter very efficiently. They generate enormous added value by drawing on complex models stored in a database or in a cloud. These values create setpoints for designing the process, and the installed CPSS

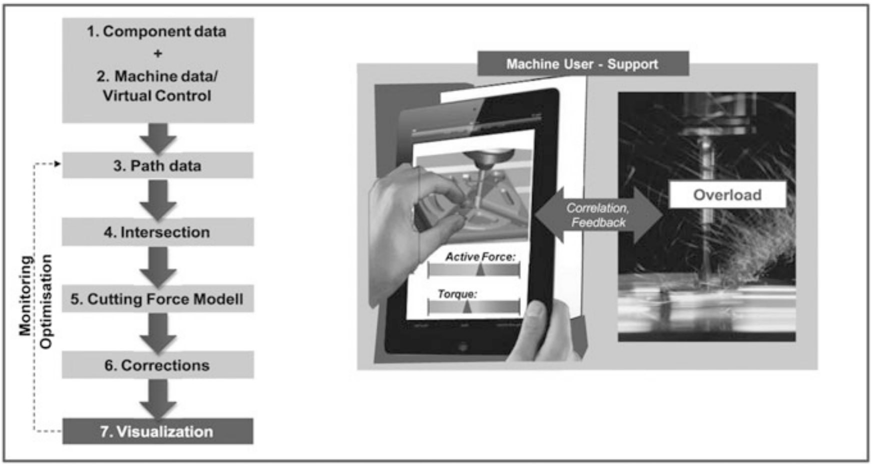


Fig. 11 Application in the production environment—Determining the theoretical cutting force [5]

sensors can compare these with the data suplied by integrated its sensors. With a suitable visualization app, this comparision can be transferred directly to the worker who will be able to exploit this information and turn it into further technology knowledge. This permits the machine operator or other entities involved in production to determine whether the manufacturing system is currently operating within the required tolerances or to its full capacity. In addition to providing direct representations of complex production and manufacturing processes, tech apps can be used to feed relevant results and new characteristic values generated in a specific application into networks. This closes the loop to the use of characteristic values and other key figures, as discussed above. These networks might represent in-company communities which form global networks used to exchange empirical values and machining results with one another. It is also conceivable that companies could share independent networks of technology experts in this way, thus contributing to technological development in a diverse range of areas of production technology.

3 Summary

The Smart Factory is one of production sciences' major issues for the upcoming future. However, it is in need of autonomous, connected and intelligent systems to create and process information that helps to monitor, control and improve processes running in the factory. The implementation of such systems can be done with Cyber-Physical Production Systems that rely on another system, the Cyber-Physical Sensor System. Both systems consist of numerous elements that are tightly webbed up with each other. The main sources of "perception" for these systems are sensors in various forms and numbers that turn conventional machine tools into smart machine tools. Those smart machine tools are able to autonomously create and assess information about the processes running on themselves, or about their own condition. Furthermore, this information can be used for self-optimization or the creation of technology knowledge. The sensors in use are constantly increasing, either in terms of adding more measurands or in the availability and affordability, making it possible to integrate numerous sensors into one system. The concept of intelligent sensor systems is currently being developed to broad extent, and many examples are presented in this work. They range from specific monitoring solutions over concepts that draw information from basic sources, such as machine internal data. But not only sensor data play a major role in the gathering and processing of data, yet also data acquired from simulation processes. These data needs to be merged with the sensor data to receive a holistic set of information. When looking at sensor data, the object from which information is drawn is not only the process itself. Workpiece orientated monitoring is especially gaining for safety-critical parts, such as aero-engine parts. And beside the very process itself, all supporting activities and tools, such as consumables, have to be considered to get the big picture of a manufacturing process. Once the data is gathered and processed, its utilization brings up new questions. These questions mainly concern the cost of information, especially in terms of computational power requirements. Model-based approaches give close and useful approximations to the real world while tremendously reducing calculation times, and cloud computing and shared services as well as worldwide accessible technology data bases build up a sound foundation on which such calculations may be held. Once the information is processed and utilized, it needs to be brought to either a system, or more likely, to a human who is able to make best use of it. This might be achieved through the use of technology apps that represent powerful means of building up interfaces between the system and humans. They can help to create knowledge, to support with a decision or assist the worker at his daily tasks. This creates coherent production systems that are there to bring our production facilities to match with tomorrows needs of increasing complexity, an aging work force and ever faster changes of the environment—the truly smart factory!

As stated in the beginning of this section, the future of industrial production is defined by the answers to a series of questions. Although industrial production is currently in a very volatile and changing situation, this section gave some answers to these questions. In order to establish semantic connections between production models in an intercompany or interdisciplinary environment, standardized communication models that feature semantic language should be used. One such protocol is OPC UA, which is open-source and able to machine-read with semantic context. Such connections would be able to gather much information about the production processes in a company, thus reducing complexity and hence giving control over global supply chains. By doing so, a lot of data will be created, for a huge variation of situations. Just like raw oil is of high value when refined properly, these data also bear a huge potential if suitable means of “refining” are found and used. One such refining aspect is meta data. Without sufficient information about the data, they will never give its user the full benefit and value that they bear within them. New businesses will emerge to provide services and tools to make use of the data, to refine them and to draw the crucial information out of the data, to harvest their value. Models will play a large role in this, as they allow merging cognitive knowledge with the information that lies within Big Data. In order to handle this large amount of data, it is necessary to process the raw data at it source of origin, and only transfer the densified information. CP(P)S play an important role in this, as they enable a sensor to already process its data right after gathering it, and by doing so reducing the amount of data transferred, and also to increase reaction times and swapping the decision to a decentralized system. However, decentralized systems always create security threats, as the number of interfaces to the outside world is increase tremendously. Hence, security aspects need to be on focus when developing CPPS. However, even the best architectures of CPPS and the best business ideas will not reap monetary benefit if they stay what they are: Good ideas. It is therefore crucial for enterprises and companies to take these ideas and implement them in their production lines.

Acknowledgments The authors would like to thank the German Research Foundation DFG for the support of the depicted research within the Cluster of Excellence “Integrative Production Technology for High-Wage Countries”.

References

1. Acatech—Deutsche Akademie der Technikwissenschaften e.V. (Hrsg.) (2013) Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0. Abschlussbericht des Arbeitskreises Industrie 4.0
2. Hardjosuwito AF (2013) Vorhersage des lokalen Werkzeugstandweges und der Werkstückstandmenge beim Kegelradfräsen. Dissertation, RWTH Aachen. Apprimus Verlag, Aachen
3. Herzhoff S (2013) Werkzeugverschleiß bei mehrflankiger Spanbildung, Dissertation, RWTH Aachen. Apprimus Verlag, Aachen
4. Hinduja S, Kunieda M (2013) Modelling of ECM and EDM processes. CIRP Ann Manuf Technol 62:775–797
5. Klocke F (2014) Technologiewissen für die digitale Produktion. In: Brecher C, Klocke F, Schmitt R, Schuh G (Hrsg.) Integrative Produktion, Industrie 4.0: Aachener Perspektiven (Aachener Werkzeugmaschinenkolloquium 2014). Shaker Verlag, Aachen, pp 247–270
6. Klocke F, Zeis M, Harst S, Herrig T, Klink, A (2013) Analysis of the simulation accuracy of electrochemical machining processes based on the integration level of different physical effects. In: Schubert A, Hackert-Oschätzchen M (Hrsg.) NSECT 2013—international symposium on electrochemical machining technology, November 12–13, 2013, Scripts Precision and Microproduction Engineering 7, Fraunhofer IWU Chemnitz, pp 165–170
7. Klocke F, Joseph Y, Trächtler A, Adams O, Backmeyer M, Blattner M, Brockmann M, Eisenblätter G, Gierlings S, Henke C, Jamal R, Kamps S, Keitzel G, Schulz K, Veselovac D, Wirtz G (2014) Sensoren für die digitale Produktion. In: Brecher C, Klocke F, Schmitt R, Schuh G (Hrsg.) Integrative Produktion, Industrie 4.0: Aachener Perspektiven (Aachener Werkzeugmaschinenkolloquium 2014). Shaker Verlag, Aachen, pp 271–298

8. Lee K (2005) IEEE 1451: empowering the smart sensor revolution sensors. In: Conference/Expo 2005, Chicago, IL June 7, pp 1022–102
9. Langmann R, Stiller M (2015) Industrial cloud—status und ausblick. HMD Praxis der Wirtschaftsinformatik 52(5):647–664
10. Pitsch M (2014) Ganzheitliche Entwicklung von Industrie-Applikationen, Doktorvortrag, RWTH Aachen
11. Rudolph T (2014) Adaptierbare Parametrierung von Diagnosesystemen auf Basis digitaler Antriebssignale zur Anwendung in der Prozessüberwachung, Dissertation, RWTH Aachen. Apprimus Verlag (Aachen)
12. Uhlmann E, Eßmann J, Winterling J-H (2012) Design- and control concept for compliant machine tools based on controller integrated models. CIRP Ann Manuf Technol 61 (1):347–350
13. Werkzeugmaschinenlabor WZL. http://www.wzl.rwth-aachen.de

Author Biography

Sven Goetz born 27.08.20187 in Neuss/Germany has studied mechanical engineering in a bachelor programme at RWTH Aachen until 2011. From 2011 until 2012, he studied production science in a master programme, and from 2011 until 2014 he studied economics in a master programme, both at RWTH Aachen University. Since 2014, he works as a research assistant at the Werkzeugmaschinenlabor der RWTH Aachen University (WZL) in the department for process and product monitoring. His fields of research are digital production, Industrie 4.0 and intelligent tool monitoring systems for milling and hobbing.

Manufacturing Cyber-Physical Systems (Industrial Internet of Things)

Ulrich Berger, Jürgen Selka, Alexandros Ampatzopoulos and Janny Klabuhn

1 Introduction

In Germany large quantities of crude oil are processed into oil and petrochemical products. The transport of the hazardous substances frequently takes place on the railways. While the production processes of liquid petrochemical products are automated to a large extent, the so-called loading into tank wagons as seen in Fig. 1 takes place manually. These activities carry a high health risk for the personnel involved, as the materials to be loaded are harmful because of their toxic properties and their high risk for explosion.

In addition, the activity of filling or loading the tank wagons can sometimes take place under harsh environmental conditions with temperatures ranging from −20 °C to +35 °C. To reduce the risk and the physical strain for the personnel, such as heavy lifting or non ergonomic working posture due to the poor accessibility of the dome cover on the tank wagon, automation of loading operations will be sought. The high heterogeneity of tank wagons available, the large varying ambient conditions and the list of statutory explosion protection guidelines present additional challenges. Due to the high risk potential of hazardous substances that are to be loaded, comprehensive security actions have to be respected to protect the

U. Berger (✉) · J. Selka · A. Ampatzopoulos · J. Klabuhn
Chair of Automation, Brandenburg University of Technology Cottbus-Senftenberg, Siemens-Halske-Ring 14, 03046 Cottbus, Germany
e-mail: berger.u@b-tu.de

J. Selka
e-mail: juergen.selka@b-tu.de

A. Ampatzopoulos
e-mail: Alexandros.Ampatzopoulos@b-tu.de

J. Klabuhn
e-mail: janny.klabuhn@b-tu.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_16

Fig. 1 Tank wagon for the transport of liquid petrochemical products [14]

environment as well as the personnel. For the reasons mentioned above, a system for automatic loading has to be integrated to the already existing automated complete system.

The solution presented determines the essential geometric parameters specifically. A purposeful integration of the plant operator secures the process flow of the system. An integrated database helps to reduce necessary operator interventions to a minimum after a learning phase. Through several stages of simulations and prototypes, a solution has been developed and validated. The use in production environment is in preparation.

2 Preconditions and Standards

The integration of CPS in machinery and equipment can mean a great revolution of the process. This represents basic and long-term changes that can only be gradually realized until the full integration is established. In an existing large-scale production environment with a high degree of automation such comprehensive conversions are usually not possible. Rather, the new system has to be integrated into the existing components. This integration usually takes place during operation. Thus, the system planning involves not only the project management and process planning, but has to be coordinated with the management of production safety of the IT department and the planning process of the entire system. The process represents a high capital expenditure that has to be calculated and planned in detail by the company's management.

Moreover, in this particular case, it is a plant with potential hazard for humans and the environment. For this reason, various laws and regulations pertaining to licensing of hazardous installations have to be considered. Important foundations for this are described in the "Law on protection against harmful environmental impacts […]" [2]. For this reason, certain partial steps must be previously checked by the safety engineer and also have to be approved by the company's management.

Furthermore, the planned system has to be checked in terms of safety and environmental protection prior to commissioning.

In large facilities, expansions take place at longer intervals. For this reason, standardized interfaces are necessary to ensure a safe integration. In collaboration with the personnel, problems can be solved independently. The goal of modern production planning is a decentralized network of several individual CPS's, which will be able to describe and configure themselves.

The opening process of the dome cover represents the beginning of a process to be automated system and thus only a part of it. The goal is the automation of the complete filling with the support of intelligent IT.

The integration requires a flexible, cost effective and reliable construction and connection technology, which allows the retrofitting of the system into existing systems. It aims to provide a tailor-made and optimized conditioning system that can be realized with the technical possibilities available on the market.

Furthermore the implementation of standards has to be respected in order to meet certain quality standards which make the machinery reliably and protect personnel at work. For the construction of an automated tank wagon filling in this case the explosion protection according to ATEX directive [4–6] must be observed, as well as relevant standards, for example, the EC Machinery Directives.

The ATEX Directive [1] divides the system into zones of hazardous areas, which forms the basis for the selection and classification of systems and equipment, including installation. The zone corresponding to 0 is the area with the greatest hazard. In zone 1, occurrence of an explosive atmosphere is occasionally expected by air mixtures during normal operation. Zone 2 encloses the entire system. Basically it can be assumed that in areas in which products or residual products may be located, explosive atmosphere can also be found. The limitation of the zones depends on the degree of ventilation. In many cases, the zone boundary to the source can be set within a radius of 1 to 3 m of gases and vapors. The inside of the tank wagon therefore is marked with zone 0.

The explosion-proof equipment, which is used in different zones, has to comply with the explosion protection guidelines from the ATEX Directive [1] intended for use in potentially explosive atmospheres. The selection of these devices has to correspond with the requirements of device group IIA (IIB or IIC) and temperature class T3 (gasolines, diesel fuels, aviation fuels).

Besides the choice of equipment, the IEC is responsible for installing electrical equipment in explosive gas atmospheres [6] or VDE 0165-1 has to be applied. This part of the standard contains the specific requirements for planning, selecting and setting up electrical installations in potentially explosive atmospheres. For the actual planning and implementation of the system for automated tank wagon filling, it should be considerated to place the electrical equipment outside the danger zones.

3 Challenges

This chapter first summarizes the global demands of the previous chapters. Subsequently, the process flow of the entire manual filling is presented. This section will be broken down to functions and described in several features. Through usage of these functions, the required process data is derived and the requirements for the automation solution are determined.

The previous sections already submit several global requirements. The explosion protection has to be maintained on the basis of liquid petrochemical products. The equipment is located in the outside area which is why it must be maintained to ensure valid regulations of environmental protection [8]. In addition, the tank wagon is exposed to weather influences, so that the temperature range of −20 to +40 °C has to be covered. Subsequently, safety requirements have to be considered. The integration of the system to be implemented requires standardized interfaces and protocols. These further requirements result in usage of compatible and hardware components which are already in use in order to minimize the need for maintenance. The requirements mentioned above are taken into account during development.

3.1 Steps During the Tank Wagon Loading

The actual filling process is divided into the following sections:

- positioning of the tank wagon,
- visual inspection of the tank wagon and the display panel,
- opening the tank wagon dome by loosening the tommy screws and visually inspecting the lid seal and the tank wagon interior,
- complete emptying of the tank wagon and the drain pipe to control the bottom valve,
- comparing the data of the tank wagon with those of the host computer of loading,
- positioning of the filler pipe,
- filling the wagons with the product,
- closing the carriage cover and the tommy screws,
- controlling the valves when tank wagon is fully loaded and finally closing the cap.

The tank wagons are transported and positioned by a shunter to the loading facility so that they can be reached with a push transport system. This transport system can accommodate up to 24 wagons. 24 wagons are filled on two tracks per layer. The filling process is controlled from an observation point that is arranged between the two tracks. Therein the filling tube and the push transport system are controlled via an operator panel. In addition, a computer program is installed, which allows the control of the filling process and keeps an accurate record of the goods loaded.

The task of filling the tank wagon is divided into two parts. One part is compromised by the activities on top of the filling plant and the other part by the activities at the bottom of the filling plant.

3.2 Activities on Top of the Filling Plant

The activities on the top of the tank wagon are executed in the order described below. The movement of the tank wagon is accomplished during the time of filling. The entire loading process is subordinated to the filling. The positioning of the tank wagon is controlled via the control room and a thrust transport system which is moved on a cable system. The positioning of the tank wagon can be done with an accuracy of about 300 mm. The transport capacity is 12 full tank wagons.

Opening the dome cover is done manually by a worker. The worker has to reach the dome cover over a side-mounted running board or a safety frame that is moved to the position of the cover. This position can vary due to different assortment of trains. Then, the manhole cover is opened and folded back. The worker may be exposed to strong weather conditions which can result in slippery or for frozen surfaces and limited visibility conditions and thus increase the hazard potential.

During the visual inspection, the seal positioned between dome and manhole cover is carefully observed. What the worker is looking for are signs of aging, such as rust, and damaged areas. Also the inside of the vessel is checked for foreign materials. If damage on the seal or foreign materials are detected, the tank wagon will not be filled.

The filling of the tank wagon is divided into several steps. The loaded tank wagons have to be positioned on the scale below the fill tube. This is followed by an adjustment of the loading system, the tank wagon and the product to be loaded. The operations of the transport system are carried out on the control panel. The time it takes to fill the tank wagons depends on the volume and the product to be loaded. Further it should be noted that both weight and volume constraints are limited to about 90 % of the capacity and therefore have to be respected. Excluding the shunting and the positioning, an approximate time of 10–15 min can be estimated for the filling of the tank wagon.

The closing of the dome cover is very similar to the opening process. Yet again, the dome is reached above the safety frame. The lid is closed and a safety check-up is performed. Consequently the shutter will be closed and tightened.

3.3 Activities on the Underside of the Filling Plant

The activities at the bottom of the filling plant will be held on ground level. The safety signs on the tank wagons have to be checked. These include e.g. the warning signs for the conveyed product and the hazard class. Likewise, it is necessary to check the tank wagons for damage and to note these in the wagon accompanying guidebook and report any issues that need fixing. To ensure that no product leaks out during transport, the correct function of the bottom base valve has to be ensured after bottling and the shut-off valve has to be closed. Should a malfunction be detected, it is necessary to empty the tank wagons, for which corresponding devices are available.

3.4 Opening of the Tank Wagon

After an analysis of the steps described above, the opening of the manhole cover is the most critical step assigned due to an increased risk of accidents. An overpressure in the boiler may cause a popping of the cover. Furthermore, in low temperatures the personnel has to be aware of the slippery surfaces on conductors and kicks which is critical with regard to ergonomics as well.

The activities for opening the tank wagon have to be executed in a bent position and taking into account the weight of the cover, which is around 25 kg. The tasks in the lower area of the tank wagon can be regarded as less critical. This can be automated in a further step. For this reason, after describing the entire loading process, the focus is placed on the activities and requirements for opening the tank-wagon.

The tank wagon is filled through the dome, which is located on top of the tank wagon. This is described in DIN EN 12561-6 [3] as a manhole cover. In reality, however, there is a great heterogeneity of different tank wagons. These differ in the design of the locks. This is partly due to the long lifetime and use of rail tank wagon and the associated repairs and modifications over time. Models constructed from 1970 till today are made from different manufacturers. An adaptation of the tank wagons cannot be realized with these models, as these are provided by train manufacturer companies. For this reason, an in-depth analysis of the opening process and a statistical study of process-critical characteristics of the tank wagons used has been accomplished. The result of this analysis was the division of the dome opening in three variant classes: type A, B and C.

The most common manhole cover complies with DIN EN 12561-6 [3] which is compromised of four tommy screws. This is the most popular manhole cover and for the sake of this example will be called type A. Type B manhole cover can be summed up as one with special fittings and screwing. Finally, type C can be summed up as all the remaining cover types that cannot be associated with the type A and type B. A manhole cover usually has an opening weight ranging from 15 to 30 kg. Variations in temperature can cause negative pressure or overpressure in the tank wagon. It is also possible that the cover glues to the tank wagon due to the former loaded product e.g. bitumen. This can lead to additional opening forces of about 500 N.

Type A and type B have the same closing mechanism but with differing positions of the tommy screws can be covered with the same automated solution which covers approximately 85 % of all manhole covers. To open the manhole cover the tommy screws have to be eased and afterwards the cover is opened longitudinally. Some fittings are not detected and therefore it has to be assumed in principle that these fittings are available. Consequently, in case of such a malfunction, the workers have to screw or unscrew the tommy screw manually in all weather conditions and non-ergonomic postures.

4 Requirements from the Manual Process to the Automation

Due to the positioning of the tank wagons, there is a fluctuating position of the manhole cover in relation to the opening system. There are also varying lengths of rail tank wagons. For automation it is necessary to locate the exact position of rail tank wagons and the manhole cover. It is also shown that different locking variations exist and need to be covered through the automated solution. These variants have to be distinguished or identified to be able to perform the opening process. In addition, relevant process parameters in the process itself have to be determined, as well as the size and position of the tommy screws. The height of the required tightening torques and forces to open the cover must be considered in the design and choice of components.

In summary, this results in the following factors that have to be observed and individually considered before development:

- explosion protection,
- extended temperature range,
- weather conditions,
- integration into the existing system,
- security,
- standardization interfaces and hardware.

The considered process has shown the following functions that have to be implemented:

- locating the tank wagon,
- locating the position of the manhole cover,
- identifying the closure system,
- locating the tommy screws,
- positioning of the tool to open the dome cover,
- opening the tommy screws,
- opening the dome cover,
- securing the operating range for protective reasons.

5 Solution Concept (Available Technologies, Intelligent IT, CPS Development)

The process flow of opening the dome cover described above is strongly influenced by the heterogeneity of the tank wagons and the manhole covers. The realized system has to be capable of dealing with various geometries of the locking mechanism. A rigid automation in the conventional sense is not able to meet these needs in this situation. There is a need for an adaptive system that is able to deal with modified geometries, thereby ensuring a stable and secure operation of the process. For this purpose, data processing and data storage has to be implemented in a corresponding manner.

Based on the process flow in the following chapter, the solution concept will be gradually introduced. First it deals with the concrete solution to the opening of the manhole cover to then explain the link with peripheral systems, as well as the control structure.

5.1 Structure of the System

In order to meet the requirements of integration into the complete system the solution concept is developed as an independent module. This will provide the flexible implementation in the overall process. Cyber-Physical-Systems provide this flexibility through open and standardized interfaces. These interfaces enable the integration of databases and external services. For a practical implementation, the characteristics of CPS are transferred to the automation solution. For this function modules can be defined as the complete closed systems with assigned task each from the process flow. The tasks are taken from the previously developed process flow. The main tasks are:

- locating the tank wagon,
- locating the position of the manhole cover,
- identifying the closure,
- locating the tommy screws,
- positioning the tool for opening the dome cover,
- unscrewing the tommy screws,
- opening the dome cover,
- visually inspecting the seal of the manhole cover and
- securing the workspace (protective function).

In addition to these main tasks, further secondary tasks have to be performed, but will be further described at a later stage. With the features listed above, the solution concept is developed. The solution concept is made up of four subsystems together are described below.

5.2 Process of the Automatic Opening of the Dome Cover

The automatic opening process of the manhole cover begins with the opening of the loading facility. Firstly, the tank wagons identity is supplied by the loading system and if not yet registered, entered into the database of the opening system. In the following step, the necessary data is read out, determined or programmed, depending on the qualifications of the operator. If it is determined in this process that the robot is not able to open the manhole cover, it is driven into rest positions as well as the linear axis and the process is completed. If the robot can open the cover and the safety measures are in working order and allow movement, the search routine for the beginning of the tank wagon beginning. Being found, the robot can be positioned on the manhole cover. After the rough determination of the position of the tommy screws these will be examined separately and measured precisely in their position and orientation. If the handling positions of the tommy screws and the cover opening routine are known and verified by the operator, the opening and folding of the tommy screws starts.

For the unscrewing of the tommy screws a certain procedure has to be considered to allow defined pressure equalization. Overpressure can escape by unscrewing the first three screws. The time for pressure equalization is estimated about 20–30 s depending on how high the pressure difference and size of the unscrewing pressure equalization is. The screwed fourth tommy screw prevents an impact of the dome cover. By tightening of that fourth screw, a vacuum can be compensated. The manhole cover, which has a degree of flexibility in the joint, thereby tilts and air can flow. Then the last tommy screw can be unscrewed.

The threads of the tommy screws are different, so for each type of screw, a number of revolutions have to be determined. Because each tommy screw is secured by a thickening at the end of the threaded bolt against turn off, the toggle can be untwisted until the end is reached. By monitoring the rotational movement and the number of revolutions the end can be determined to open the knob and damage the tank wagon or the mechanism can be prevented. With this information the tommy screws can be opened.

When all tommy screws are unscrewed and folded down, the dome cover can be opened. For this purpose, a pneumatically operated permanent magnet is at a predefined point, which has a surface as large and flat as possible to lift the manhole cover. This permanent magnet can be moved in its position by compressed air and thus switch the magnetic effect at the contact surface to the cover on and off. The magnet must be flexibly mounted to the tolerances of the curvature of the manhole cover and to compensate for the potential error of the aperture radius. After reaching a tilt angle of about 95°, the magnet should be released and the manhole cover should be stored slowly over a defined and padded on the gripper or by a holder to avoid an impact of the manhole cover on the tank wagon. It must be noted that the opening angle of the dome cover can be up to 180°.

Subsequently, the seal of the manhole cover has to be checked, therefore it is necessary to take pictures of the manhole cover and the manhole opening.

Finally, the interior of the tank wagon is documented. This allows the operator to carry out a review of the level and a check for foreign objects inside.

5.3 Vision System

For the problem of the fully automated manhole cover opening, a large number of functions for geometry recognition are necessary. For this reason, the use of machine vision is particularly suitable. One advantage is that the vision system can perform multiple tasks. A second advantage is the ability to send important data to the viewpoint and therefore to the monitoring personnel. Thus, they have a Life-control in an optimal form for humans.

For industrial processing, there are sensors that detect 2-dimensional and 3-dimensional objects. The advantage of 2-dimensional image recognition is the large availability and sophisticated technology of the RGB sensors. A disadvantage is the dependence between the light and the picture quality. A method for 3-dimensional component recognition may represent low image sensors, stereo camera systems and laser scanners. Depth image sensors and laser scanners work in a specially optimized workspace. Thus, a special sensor must be selected for each individual task. These RGB sensors however, can work regardless of the lighting situation and are independent of interference sources. During the subsequent processing of the image data, the depth information is determined from the offset.

A combination of both types is the Kinect camera system. This is because the Kinect camera system uses a patented method which is published under the name "Light Coding" [13]. The central elements of Kinect are an RGB camera, an infrared (IR) -depth sensor and a microphone array [11] . With the use of this RGB camera 2-dimensional images are produced. The Kinect camera depth sensor consists of an infrared projector and an IR camera [13]. The projector continuously transmits light having a wavelength in the infrared region (PrimeSense ™ 3D sensor). This light is invisible and harmless to humans. The Kinect camera system can also be easily equipped for use in hazardous areas.

This system can be used for all localization applications. This system allows for multiple uses of the programming effort and the number of required hardware components decreased which in turn leads to lower costs. Furthermore, only one sensor area must be equipped for use in the hazardous area. From the Kinect camera is the vision system in conjunction with an industrial PC for image evaluation.

The localization of the tank wagon is realized when the front part of the tank car is determined. To determine the front of the tank wagon, the edge of the tank wagon has to be recognized by the 3D camera as depth image information. With the help of the RGB camera, the nameplate is read out with the wagon identity and compared to the database. This serves to determine the most probable position of the dome cover from the data of the location of the tank car. The manhole cover itself is detected by means of the circular geometry. The alignment of the lid can be detected based on the geometric features of the tank wagon.

These features may turn out to be the joint of the dome cover or the handle for opening of the lid. Furthermore, additional structures can be used to identify the manhole. These marked points must be programmed separately for the different types of tank wagons. The algorithm cannot be identical for all tank wagons because of the large number of different variations. For each new type of tank wagon, the appropriate peculiarity and the location to the center of the dome cover must be taught in.

The investigations have shown that the depth sensor of the Kinect is very well suited for detecting the location and identifying the dome cover. A unique detection is possible for well known cover versions in the system. After the position and the type were determined each tommy screw is measured individually using the depth sensor in a subsequent step. By individual measuring the exact orientation and the point to open the tommy screws is determined.

A statistical data collection could be made to classify the types of closures. The manhole covers are classified according to their characteristics into three groups type A, B and C. This procedure is used for the rapid detection of closure type of the manhole cover. The lid variant of type A is characterized by four thumbscrews. Type A cover variant is the most commonly encountered design with about 75 %. Type B is characterized by a bracket which extends over the manhole cover and is secured with a screw. The strap is firmly attached to the dome cover and aligned along the direction of travel. Cover variants of this group account for about 10 %. Type C includes all the rest of the cover versions that are not respectively associated with the types A and B. These include, for example, complex variants of combinations of tommy screws and strap closures. By classifying, it is possible for the specific identification process to search for existing geometries. These geometries can be identified through the depth image information. This makes it possible to assign the present dome lid to the corresponding type.

The control of the dome cover seals cannot be performed using the depth sensor. This is because the resolution is too low to be able to detect defects such as cracks or dirt. The depth image sensor can only detect objects being greater than 24 mm at distance. Flaws that are smaller than this minimum size cannot be reliably identified with the depth sensor. The data of the color image are very dependent on the ambient lighting conditions, the color and texture of the seal and therefore difficult to evaluate.

A combination of both sensors color and depth information promises the best results. The position and location of the seal is acquired with the depth sensor. In this case a color image with a higher resolution has to be searched for as the depth image is looking for shade or hard transitions during sealing. Hard transitions and shadows are identification criteria which can provide evidence of flaws. This effect can be enhanced with appropriate lighting. The results can then be made available to the plant personnel for evaluation. The color of the camera with its relatively low resolution has only limited suitability examined through experiments. Good results were achieved with frames that show a close-up shot of the seal.

As stated above, the localization of the tank wagon, the dome cover and the tommy screws are covered by one single vision system that simultaneously takes the task of identification of the cover type. This vision system uses a 3D camera with depth imaging and an RGB sensor. It recognizes unique geometries and positions of the different closure types of the dome cover. The camera system can be adapted to the required environmental conditions. One difficulty which has not yet been satisfactorily solved, is the visual inspection of dome lid opening seal. The low resolution of the RGB sensor of the Kinect can be compensated by close-ups shots. The widely varying lighting conditions outdoors may prevent reliable detection.

Nevertheless, by using a single camera system results in a further required function. The camera has to be repositioned depending on the process step. The function of positioning is already present in the main tasks and is described later in the appropriate section.

5.4 Gripper System

Through the vision system, the position data and the type of closure are known. This data is needed to perform the opening process. First the tommy screws have to be unscrewed depending on the type the corresponding shutter. For this purpose, a special gripper system must be developed that is capable to open different closure types. It is also intended to open the dome cover through the developed gripper system. This process is generated by a pneumatic rotary drive. The torque is transferred to the tommy screws through a tool adapted to the geometry.

At the time of the first validation of the opening process, the tool is initially only able to solve tommy screws. A further development is planned. The opening of the dome cover is realized with a permanent magnet. This is also pneumatically adjustable and has the advantage that no handle or similar device on the dome cover must be present to perform the opening process. The permanent magnet has to be placed on a nearly flat spot on the dome cover. This results in the additional task of positioning the magnet as well as the tool to open the tommy screws. The following section describes the positioning.

5.5 Robot System

The positioning of the tool to open the dome cover is a major task. In addition, the vision system must be, as previously explained, also moved and positioned. The requirements for the position are a flexible orientation in all six degrees of freedom. These include the three translational degrees of freedom and three rotational degrees of freedom. This results from the different variants of the dome cover and the orientation of the tommy screws. For this reason, a 6-axis industrial robot is used. This enables free programming of six degrees of freedom. To provide additional workspace to the robot it is mounted on a linear axis. This allows the robot to shut down the tank wagon along the entire length. This allows the vision system to determine the beginning of the tank wagon.

The most important requirements for the robot are initially reach, load and the torque of the last axis. To determine the necessary reach of the robot according to the loading gauge, railway construction and operating regulations are analyzed [9]. The loading gauge describes the maximum dimensions of rail vehicles. The robot must be outside this range, so that a collision with the tank wagon cannot occur. Nevertheless, the robot must be positioned close enough to reach the dome cover. Analyzing the loading gauge a tank wagon is more than 3150 mm wide. The horizontal distance of the dome cover to the robot is essential for the operation of the robot. The working area of the robot should be large, central and sufficiently deep enough over the manhole cover due to the different levels of tank wagons.

In addition, the robot must be equipped with an ATEX certification and be approved for the extended temperature range. For ATEX certification two alternatives are available to choose:

- A *sleeve* can be used. This can be carried out simultaneously conditioned to meet the requirements to the ATEX certification and the extended temperature range
- *Conversion solutions* for the robot are exiting, so that no additional cover is needed.

A shell must be able to take the necessary cables and extensions, and may not restrict the freedom of movement of the robot. The alternative is the conversion of the robot for the ATEX [1] certification and to build a shelter for the working area, so that the robot is not exposed to direct influence by snow, rain and sun. The actuator must be ATEX-certified and designed for the extended temperature range. By choosing a robotic system of ATEX-certified and for the extended temperature range operational 6-axis industrial robot and a linear axis, the tasks of positioning of the vision system and the gripper system are therefore fully met.

5.6 Security System

This section describes the identified risks arising from the operation of automatic dome lid opening system for the operator and for the workpiece. Furthermore, additional technical protection measures to be are explained.

5.6.1 Risks

A principle of robotic automation in the production says that robots and humans cannot meet. Dangers that may arise from the automatic dome lid opening system are for humans often completely unpredictable and occur in complex patterns of movement, rapid speed changes and at the same time from enormous forces released. Work alongside an unsecured industrial robot can be rapidly fatal for a human being. For these reasons, industrial robots almost always work behind safety fences, screened from the application operator. The aim is to avoid collisions.

Although new security controls can prevent the robot to leave a programmed workspace, the entry of the employee in the workspace of the robot cannot be excluded. In this context, all possible phases of operation of the automatic dome lid opening system should be considered: the normal operation of the system in automatic mode, as well as repair and maintenance work in manual mode. In this consideration, it must be examined which operations are executed by people and which by the machine.

In automatic mode, no personnel may be located directly in the system, so here the work is executed by the machine. An exception is the repair and maintenance work, in which people are located in the plant. Another operating condition is the state in which the robot for various reasons does not perform its intended function. Fails can be failure of components, external interference, design flaws and imperfections and faults in the power supply and ambient conditions.

5.6.2 Technical and Complementary Protective Actions

In designing the machine the operator must be able to ensure the main steering position, so that no person is exposed to the hazard zones [3]. Therefore, an adequate view from the viewpoint must be ensured in the direction of the automatic dome lid opening system. This is made possible by a complete window front, as well as video surveillance at the facility.

The mechanical components of the robot are arranged so that no hazardous actions can occur by crushing or shearing. The robot is observed beyond a certain radius. Furthermore, the robot is limited in its speed and movement through a secure robot control and further acceleration or deceleration and motion limiters are fitted. The electrical equipment is appropriate to the IEC 60079-11 [7] selected and is considered intrinsically safe. Furthermore, the robot is designed to be stable and incorporated into the automatic dome lid opening system.

Hazards and the associated risks can be mitigated by an appropriate choice of design features of the robot and the system. Despite previously met security measures, it is still necessary to construct a fixed guard action, e.g. a protection fence, in the workspace of the robot. The safety fence can be removed only with the aid of a tool and is there to block the workspace from the robot attached to the front of the dome cover. This is intended to give protection e.g. against catapulted parts.

Furthermore, over the tank wagon, a non-separating precaution is planned called light curtain. This light curtain ensures that no personnel enter the area of the robot, by rising to the side or middle ladders fixed on the wagons. The light curtain shown in Fig. 2, surrounds the dome cover and is disabled during the automated opening process to avoid failures caused by bad grasp of the robot in the manhole cover area. Tracking the tank wagon from the plant by the shunter, the light curtain is active. Another area outside the automated dome lid opening area is reserved for rework. Here all the light grids have to be turned on to enable the workers to begin the rework.

To allow access to the plant despite a suitable safety concept, a protective door is attached with a safety lock. The precautions will ensure that unauthorized access of any person is disabled for automatic dome lid opening system. The Fig. 2 shows an array of protective measures in simulation.

To allow access to the plant despite a suitable safety concept, a protective door is attached with a safety lock.

5.7 Total System: Dome Cover Opening System

To execute the dome lid opening process, the following featured systems have to work hand in hand. The vision system and the gripper system are mounted together to one single gripper. This gripper is positioned by the robotic system during dome lid opening process and the work area is monitored and protected by the security system. The merger of the four systems results in the parent dome lid opening process.

The communication and control of the subsystems between each other is performed by a cell controller. The industrial robot with linear axis and robot controller are connected to this cell controller. Communication takes place via the Profinet interface. Furthermore, the vision system and other required actuators and sensors are connected with the cell control. Communication with the peripheral sensors and actuators is achieved via the Profibus interface. The Vision System consists of a 3D camera, an image processing and a data preparation industrial PC. The coupling of

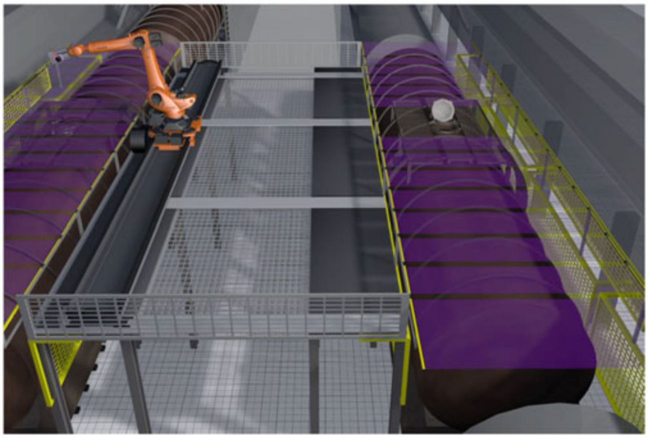


Fig. 2 Simulation of a protection system

the vision system with the cell control takes place via TCP/IP. Even within the Vision system, the 3D camera communicates with the industrial PC via TCP/IP.

Communication with higher-level systems is performed by the cell control and the industrial PC, see Fig. 3. The other relevant plant systems for the processes are described in the following section.

The particular difficulty of implementing the dome cover opening process lies in the stability of the process itself. It cannot and should not be excluded that an unknown tank wagon type must be processed in accordance with an unknown locking system. Conventional automation solutions are defined exclusively in the situation to run to execute rigid processes. The special feature of this automation solution lies in solving this problem. For this purpose, an intelligent algorithm is used with the changed boundary conditions that also allow a stable operation process. In addition, existing peripheral information of the filling via standardized interfaces are coupled with the dome lid opening. The dome lid opening system is integrated into the existing system and linked on a horizontal plane. This decentralization creates a stand-alone system that can operate independently as CPS. In order to illustrate the link to the peripheral systems, which provides the available process data which are necessary, it is important to describe the flow of the entire process. At this point an intelligent algorithm for dome lid recognition is scheduled.

The manhole cover opening system is linked to four other systems:

- Loading system,
- Transport system,
- User Interface and
- Database System.

These systems provide information on the implementation and optimization of the dome lid opening process. Figure 4 shows the links which will be described.

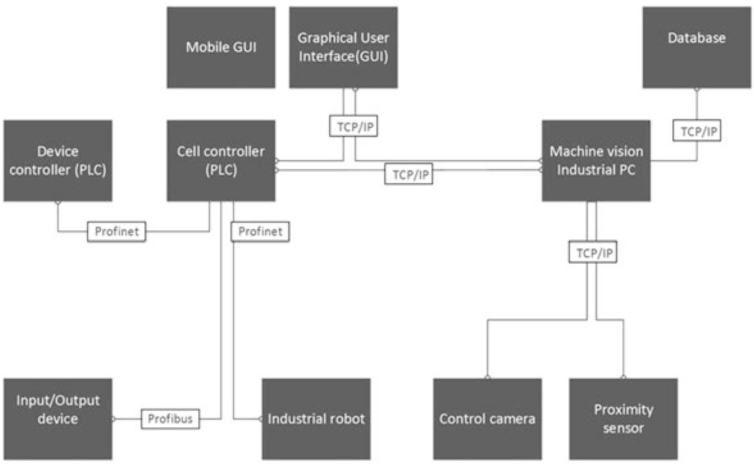


Fig. 3 Communication and integration of the port system in the whole system

Each tank wagon has its own identification number and therefore is clearly identifiable. The distribution system monitors the tank wagon number and provides it to the dome lid opening system. With each tank wagon number, an entry is generated in the database system. Through the database record, the data of the tank wagon length and type of closure for example, can be stored and is available for later use. Furthermore, analyzing the data is made possible. The transport system is responsible for the positioning of the tank wagon in the filling plant. Accordingly, the transport system features the position data of the tank wagon in the system with the help of the search algorithm to determine the position of the dome cover on the tank wagons can be accelerated. The interface forms the link to the personnel.

The Human Machine Interface is used for control purposes and allows human intervention during the process. The workspaces of humans and robots are separated. The robot is operated via a remote control station, which is not tied to the proximity to the shipment. The monitoring of the system via cameras and operation is done with a PC. For the plant personnel on site, there is a terminal available to the system where the personnel are able to control the robot from. The main operation is performed on the following screen. There, the image processing position and orientation of the tommy screws are displayed using augmented reality.

The operator can easily detect the position of the tommy screws. Should the operator spot some differences, the shown tommy screw has to be selected to open the edit screen. The programming interface is constructed so that the parameters of the algorithms may be changed. However, this does not occur when entering numerical values but by comparing sample images. Thus, several parameters can be changed simultaneously. Incorrect entries can be prevented by introducing redundancy. For new rail tank wagons or modified tank wagons there should be a programming interface, which enables parameter setting and extension of the image processing. In order to meet this great spectrum, the interface provides for different qualifications of personnel. This is done in order to distinguish between operators, programmers and administrators.

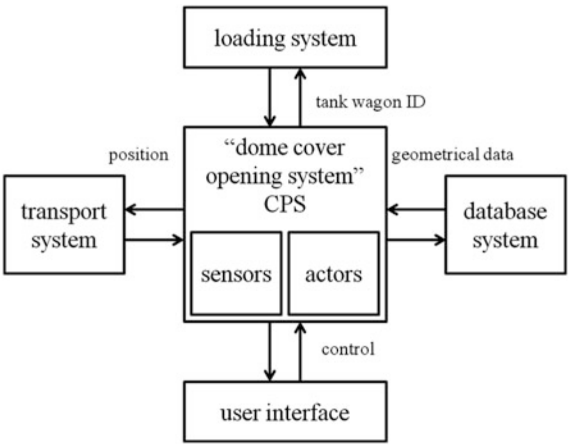


Fig. 4 Linking of the peripheral systems the dome lid opening system

The *operators* have the fewest rights. They can check whether the image processing has recognized the handling points correctly or to start or cancel a modification or an opening.

The *programmer* has the possibility to create new tank wagons, set targets, as well as select from a library created earlier and algorithms with which the handling points are determined to be parameterized.

The *administrator* creates new users and can add new algorithms in the library.

With a fixed version in the control room and a mobile version the process data can be visualized and allows an interaction with the process. The database system is used for the storage of process data from previous processes and presents them for statistical evaluations.

The tank wagon is positioned on the transport system in the plant. The exact positions of the tank wagon and the dome cover are initially unknown and therefore have to be determined. The dome lid opening system receives the information through communication with the transport system and on certain tank wagons a number from the database system. This makes it possible to detect the dome lid position using statistical methods. To open the dome cover, the exact position data of the locking system of each tank wagon is required. As shown in Fig. 4 is the tank wagon ID is determined as a first step and then the database is checked on existing data.

The implementation of the opening-up process requires an alignment of the opening tool in three-dimensional space around the tommy screws to loosen and unfold the manhole cover. To meet these requirements and combine both processes in an execution system, a 6-axis robot is used on a linear axis. Thus, the accessibility of the dome cover in different tank wagons is assured and the flexibility with respect to any new generation of tank wagons with modified geometries is increased. The used tool of the robot has two main functions in the process flow. The specifically developed gripper assembly firstly determines the process-relevant parameters using a vision system and secondly it leads with special actuators the actual physical process of opening the tommy screws and the dome cover. Before the position of the dome cover can be found the front part of the tank wagon must be known. For this purpose, the robot performs a search operation and determines the front part of the tank wagon based on depth image information of the vision system. This is done with the data from the transport system.

5.8 Flow and Operating Concept

Along with the tank wagon number and the statistical data from the database system, the most probable position of the dome cover can be determined on each tank wagon type.

In the next stage, the tank wagons are moved by the transport system into the facility. This allows the transport system to gain resulting position information of the tank wagon as shown in the flowchart in Fig. 5. In the next step the tank wagon's identity is determined by the distribution system. If the tank wagon's identity is already stored, then the available data for the next process step is provided. If the tank wagon's identity does not exist in the database system, a corresponding record is created. The following step is the process of opening the dome cover. This is to be traversed into several steps from the dome lid opening system. This process is started over the human machine interface.

The previously determined tank wagon identities are checked whether the current tank wagons can be opened by the opening system. If this is the case, the protection system is activated, which switches the light grating. After switching on the light grating, the robot can start with the first search drive. This is determined by the vision system and the depth image information of the tank wagons front. In an existing data set from the vision system parameters such as exact location of the dome cover diameter and on the tank wagons are used to speed up searching. This procedure is only possible through an adaptive search algorithm. If the tank wagons are unknown, a manhole cover-search journey is followed. Through the statistical analysis of the tank wagon records from the database, this search operation can be accelerated. To this end, the most probable position of the dome cover is calculated. Once the dome cover is found, the robot is positioned over it.

Through the vision system, the center of the dome cover and the radius can be determined. This data will be compared with the database and if not available yet they data will be stored. As a next step, the pivot point of the lid is identified and from there the axis of rotation for opening the dome cover will be determined. Before opening the lid, a starting point for the magnet of the gripper system has to be localized. Before the manhole cover can be swung open the toggle bolts are measured and displayed. For this purpose, the coarse position is determined so that each bolt can be approached and measured.

The orientation of the screw is determined and a handling point for loosening the screw is calculated. After all the screws have been surveyed, the data on the HMI is transmitted to the operator. The operator then checks the calculated points of handling and gives clearance. Should a handling point be faulty, the operator corrects this via the HMI. If all targets are correctly aligned, the operator starts the opening sequence. The control is created on the basis of the process data and information collected from the database a on each tank wagon individual opening algorithm. Thus, the robot is capable to unscrew the tommy screws sequentially and finally open the manhole cover. In the final stage, the robot is brought into a defined rest position.

6 Research Challenges

In order to open the tank wagon a cyber-physical system was developed. It represents an important module for a fully automated loading of petrochemical products. With its help, the plant personnel can work more ergonomically and in a much safer environment. CPS consists of a vision system which takes over the

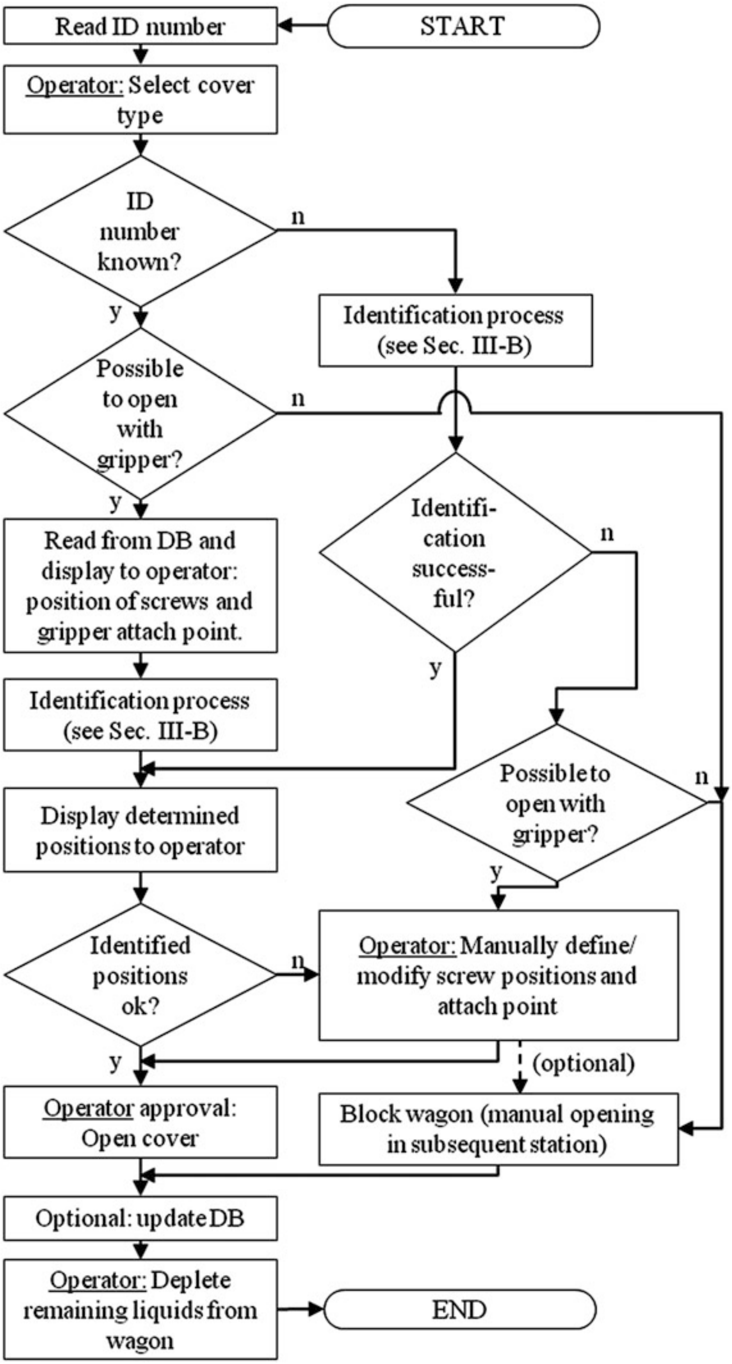


Fig. 5 Flowchart [12]

identification of the tank wagons and the localization of the geometric characteristics. A gripper system for the automatic opening of approximately 75 % manhole covers was developed. 100 % coverage is currently not possible because of the large heterogeneity of opening systems developed for humans. Furthermore, the quality control of dome lid sealing must be further investigated. Because of the uneven lighting conditions outside an automated evaluation is not possible. To transfer the captured images to the control room it is important to maintain sufficient image quality. However, dangerous situations when entering the tank wagons can be avoided.

The greatest difficult y is the operation of the industrial robot in the extreme temperatures between −20 and +30 °C. Although some protection concepts exist, they still require weatherproof capabilities. Furthermore, the robot must be protected even in rest periods against low temperatures what requires continuous heating.

7 Conclusions

The step by step transition from manual processes to automated solutions requires a long-term vision for future system expansions as well as to consider short-term solutions. There, the special requirements were pointed to the automation of a refinery and expanded to include the modern demands of the continuous integrated IT.

The development of the system rejects to a concept of cyber-physical systems. The system concept is provided in the foreground by the structure of the automation solution.

For this purpose the vision system, gripper system, robotic system and security system were summarized based on the functions to be performed and developed as automation solution. The following main features were determined first from the manual process in the current situation.

- Locating the tank wagon,
- Locating the position of the manhole cover,
- Identifying the closure,
- Locating the tommy screws,
- Positioning the tool for opening the dome cover,
- Opening the tommy screws,
- Opening the dome cover,
- Visual inspection of the seal of the manhole cover and
- Securing the workspace (protective function).

Similar functions were summarized and each assigned to one of the four systems. These cover all the main functions and form the advanced automation solution. Next, peripheral conditioning systems have been taken into account and integrated via defined interfaces to create a consistent communication structure.

This integrative approach allows optimization and acceleration of the process through an intelligent algorithm that can access external data. Embedded systems are the database system, the user interface, the docking system and the transport system. The database stores process data and allows a statistical evaluation. When it comes to the user interface, a process monitoring and control is implemented. The operator can perform adjustments of the process "online". The loading and transport system transmits tank wagon and position data to the developed automated dome lid cover system. Then the automated process flow has been shown and explained using a flow chart. By automating this process, harmful and non-ergonomic working processes could be eliminated. Due to the strong heterogeneity of the tank wagons at the current state not all manhole cover can be opened automatically. Nevertheless, this case is considered in the process and does not lead to failure of the system. The next step is to test whether the permanent deployment of the robot under climatic influences leads to increased consumption or shorter lifetime. Overall, it could be shown how the approaches of cyber-physical systems can be connected and implemented to classic automation technology.

References

1. ATEX (2014) Directive 2014/34/ EU of the european parliament and of the council of 26 February 2014 harmonization of the laws of the member states concerning equipment and protective systems intended for use in potentially explosive atmospheres
2. BImSchG (2002) Law on protection against harmful effects of air pollution, noise, vibration and similar phenomena— recast from 26 September 2002—BGBl. I, p 3830
3. DIN German Institute of Standardization e.V. (2011) Explosive atmospheres—Explosion protection—Part 1: basic concepts and methodology, EN 1127-1:2011
4. DIN German Institute of Standardization e.V. (2011) DIN EN 12561:2011: tank wagon part 6: manholes. Beuth Verlag GmbH, Berlin
5. DIN German Institute of Standardization e.V. (2011) Non-electrical equipment for use in potentially explosive atmospheres—part 5: protection by constructional safety 'c' EN13463–52011
6. DIN German Institute of Standardization e.V. (2012) Potentially explosive atmospheres—part 11: equipment protection by intrinsic safety "i"(IEC 60079-11:2011 + Cor.:2012), EN 60079-11:2012
7. DIN German Institute of Standardization e.V. (2012) Potentially explosive atmospheres—part 14: design, selection and installation of electrical systems (IEC 31 J/208/CDV:2012), FprEN 60079-14:2012
8. DIN German Institute of Standardization e.V. (1995) DIN EN 60721 Classification of environmental conditions part 3: classification of environmental parameters and their severities Section 4: Ortsfester Einsatz
9. EBO (2015) Eisenbahn-Bau- und Betriebsordnung (Railway Construction and Operating Regulations) from 8. May 1967 (BGBl. 1967 II S. 1563), Artikel 518 der Verordnung vom, 31 August 2015 (BGBl. I S. 1474)
10. Freedman (2010) Patentnr. US 2010/0118123 A1, USA
11. ifixit (2010) Microsoft Kinect Teardown. http://www.ifixit.com/Teardown/Microsoft-Kinect-Teardown/4066/1. Accessed 05 Nov 2015
12. Lehmann C, Städter JP, Berger U, Keller S, Gnorski H (2013) Semi-automated handling of manhole covers for tank wagon processing using industrial robots. In: IEEE/RSJ international conference on intelligent robots and systems (IROS), Tokyo, Japan, 3–8 November

13. PrimeSense (2013) Developers. http://www.primesense.com/developers/get-your-sensor/. Accessed 16 May 2013
14. Städter JP, Lehmann C, Halbauer M, Berger U, Gnorski H, Schulze T (2014) Automated robotic dangerous loading in explosion-protected outdoor area—technical security. Springer-VDI-Verlag, Düsseldorf, pp 40–46

Author Biographies

Professor Dr.-Ing. Ulrich Berger studied mechanical engineering at the University of Stuttgart. From 1984 to 1990 he assumed various management responsibilities in research, development and applications engineering at the Robert Bosch Group, Stuttgart. Afterwards he was appointed as technical director at the BIBA institute Bremen and from 1995 to 1998 chief engineer at the Institute of Material Science, Bremen (IWT). He received his doctorate at the University of Bremen in 1995 under the supervision of Prof. B.E. Hirsch. Title of thesis: "Automated robot systems for one-of-a-kind manufacturing". In 1998 he was appointed as professor for manufacturing systems at the Lüneburg University. In October 1st, 2001, Professor Berger was appointed as Chair of Automation Technology of the Brandenburg Technical University Cottbus. From 2003 to 2007 he held the office of the Dean in the Faculty of Mechanical, Electrical and Industrial Engineering at the BTU. In 2005 he was awarded the Science Ambassador of the State of Brandenburg. Since 2008 he also is guest professor at Denmark's Technical University (DTU) in the field of manufacturing engineering. In September 2012, he was appointed by the Minister of Economic Affairs of the Land Brandenburg to spokesman for the country's metal cluster. In April 2013 he was elected as Chairman of the National Federation of Regional Association of Berlin-Brandenburg by the Association of German Engineers (VDI). He was Chairman of the VDI district association for 6 years. Professor Berger is also a member of several technical and scientific societies in Germany and abroad.

Jürgen Selka is a scientific researcher at Brandenburg University of Technology Cottbus-Senftenberg, Germany at the Chair of Automation Technology. He holds a diploma in engineering from Brandenburg University of Technology Cottbus-Senftenberg, Germany. His fields of research include engineering, robotics, human machine interfaces and digitization. Within his current project at the Chair of Automation Technology he examines the possibilities for small and medium-sized enterprises to implement modern technologies.

Alexandros Ampatzopoulos is a researcher at Brandenburg University of Technology Cottbus-Senftenberg, Germany at the Chair of Automation Technology. He holds a BA in International Business and Globalisation at De Montfort University, U.K. an MBA in Business Administration at the Lincoln Business School, University of Lincoln, U.K.

Janny Klabuhn is a PhD student in Industrial Engineering at Brandenburg University of Technology Cottbus-Senftenberg, Germany. She holds a diploma in Industrial Engineering from Brandenburg University of Technology in Cottbus, Germany. Her fields of research include human resource management, innovation management, and automation technology. She is part of a research project at the Chair of Automation Technology that aims at the development of a service robot to improve health and safety in the workplace. Within this project, she analyzes the transformational processes in human resources originating from the increasing application of automation technology.

Cyber-Physical System Intelligence

Knowledge-Based Mobile Robot Autonomy in an Industrial Scenario

Tim Niemueller, Frederik Zwilling, Gerhard Lakemeyer, Matthias Löbach, Sebastian Reuter, Sabina Jeschke and Alexander Ferrein

1 Introduction

Industrial processes are currently undergoing major changes by introducing ever more *cyber-physical systems* (*CPS*)—which are computing elements combined with physical sensing and interaction. The *Industry 4.0* vision builds on such CPS and is characterized as revolving "around networks of manufacturing resources (manufacturing machinery, robots, conveyor and warehousing systems and production facilities) that are autonomous, capable of controlling themselves in response to

T. Niemueller (✉) · F. Zwilling · G. Lakemeyer
Knowledge-Based Systems Group, RWTH Aachen University, Aachen, Germany
e-mail: niemueller@kbsg.rwth-aachen.de

F. Zwilling
e-mail: zwilling@kbsg.rwth-aachen.de

G. Lakemeyer
e-mail: lakemeyer@kbsg.rwth-aachen.de

M. Löbach · S. Reuter · S. Jeschke
IMA/ZLW & IfU, RWTH Aachen University, Aachen, Germany
e-mail: matthias.loebach@ifu.rwth-aachen.de

S. Reuter
e-mail: sebastian.reuter@ifu.rwth-aachen.de

S. Jeschke
e-mail: sabina.jeschke@ima-zlw-ifu.rwth-aachen.de

A. Ferrein
Mobile Autonomous Systems and Cognitive Robotics Institute, Aachen University of Applied Sciences, Aachen, Germany
e-mail: ferrein@fh-aachen.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_17

different situations, self-configuring, knowledge-based, sensor-equipped and spatially dispersed and that also incorporate the relevant planning and management systems" [32]. Smart factories are considered to be a key component in this effort [32]. They are context-aware facilities in which *manufacturing steps are considered as services* that can be combined efficiently in (almost) arbitrary ways allowing for the production of various product types and variants cost-effectively even in small lot sizes, rather than the more traditional chains of mass productions facilities which produce only a small number of product types at high volumes [53].

Cyber-physical system intelligence describes the holistic system and the components—which are inherently principled, methodological, and software-driven[1]—that are required to achieve industrial process autonomy in this context. The overall field is vast with a plethora of methods from the field of artificial intelligence being applicable. In this chapter, we focus on task-level reasoning, planning, and scheduling to achieve autonomous decision making.

To describe and evaluate such methods, we describe the deployment of a group of autonomous mobile robots as CPS for in-factory logistics in the context of a smart factory as our testbed. It provides a limited form of the *supply chain optimization*[2] (*SCO*) problem, which is generally a hard problem [57] but which can be handled in a reasonably sized testbed. Robots are one of the most complex forms of CPS. While *mobile* robots are not widely deployed, yet, we expect this to change considerably in the near future, as indicated, for example, by efforts like the Tapas EU project [63]. More specifically, we build on experience gained from the RoboCup Logistics League (RCLL) as a testbed. It models a smart factory at a comprehensible scale where groups of three robots must maintain and optimize the material flow and handle processing stations. It constitutes a medium complex domain that focuses specifically on challenges regarding autonomous decision making and robust multi-robot task execution [49].

In the following Sect. 2, we further elaborate on CPS in a smart factory and the components necessary to achieve autonomy. We then overview possible methods to implement such components in Sect. 3. Section 4 introduces the RoboCup Logistics League which was used as evaluation testbed. We analyze several case studies of decision making and executing components as case studies that have been implemented the respective sub-sets of methods used in Sect. 5 with evaluation results in Sect. 6. We conclude with Sect. 7.

[1]To meet the challenges posed by these aspects in particular requires an increased cooperation and interdisciplinary teams of manufacturing industries and computer science research, an observation shared in [56].

[2]Supply chains describe logistic networks which comprise interlinked logistic actors [47]. In this chapter, we focus on intra-logistics for the material flow of a smart factory by a group of autonomous mobile robots.

2 Autonomy of Cyber-Physical Systems (CPS) in a Smart Factory

Modern production processes require physical agents that observe the environment, make decisions, and act in the real world. In an industrial context, such agents are *cyber-physical systems* that integrate computation and physical processes [37]. The term has undergone several development stages starting with simple identification devices such as RFID[3] chips, followed by devices with sensors and actuators that required centralized processing and instruction. The latest generation—considered as the canonical CPS [25]—are embedded systems that extend this by integrating computation and are usually networked. In the context of this chapter, we focus the meaning to one of the most complex versions of such CPS—*autonomous mobile robots*. Such systems have a variety of sensors and actuators to perceive, manipulate, move within the environment, and coordinate to form a multi-robot system.

We expect such CPS to operate in a *smart factory* (SF), a context-aware production facility that considers, for instance, object positions or machine statuses to assist in the execution of manufacturing tasks [40]. It can draw information from the physical environment or a virtual model, e.g., a process simulation, order, or product specification. It is designed to cope with the challenges that arise from the desire to produce highly customized goods which result in the proliferation of variants [40] and therefore smaller lot sizes. For the remainder of this chapter, we envision a SF that offers various manufacturing services, i.e., a number of machines that can refine, assemble, or modify a workpiece towards a final product. Given a product specification, the SF determines the required processing steps and involved machines. Such a SF rather offers a number of special production technologies than just production types [25] and requires more complex logistics for efficient parallel production. The challenge then becomes developing a multi-robot system capable of efficiently determining the assignment of robots to logistics tasks, and machine handling. The robots must coordinate in order to avoid resource conflicts (e.g., two workpieces requiring processing at the same machine) and cease the opportunity for cooperation (e.g., to meet a tight schedule for a high value product). This entails to study various methods for the task-level executive and coordination.

2.1 Components for CPS Autonomy

Many industrial sectors have achieved a high degree of automation in production processes. Such automatic systems will do exactly as programmed, they have no choice. Autonomous systems, however, have a choice to make, free of outside influence [8]. This means in particular, that they need to make such choices themselves, i.e., without human interference. For *CPS in a smart factory* to achieve

[3]Radio Frequency Identification, passive information retrieval via wireless communication.

autonomy, a number of concepts and components is necessary, from representing knowledge about the environment and its current state, over decision making what to do next, to robust execution and fault tolerance. For networked CPS, additionally communication and coordination is necessary to avoid or resolve resource conflicts and take steps cooperatively. In the following, we introduce the terminology used to describe such components, where CPS are considered to be a group of *autonomous mobile robots*.

Task-level Executive. To make decisions and act on its own without external instructions, a CPS requires a component that determines what action to take next, to execute that very action and to monitor that execution. This overall process is performed by the *task-level executive*. It requires sub-components to determine, execute and monitor tasks. For this, it typically needs a way to represent its foundational domain knowledge and its information available at run-time. By our definition, the task-level executive is what makes a CPS an intelligent system and provides the necessary flexibility to execute the suitable action in different situations.

A task-level executive can be implemented according to several strategies, for which we provide some characterization [1, 49]. It can have a *local scope*, i.e., it plans for a single agent, or it can consider the overall fleet at a *global scope*. It can operate in a *centralized* fashion, collecting all information in a single place and generating a plan,[4] or there can be a number of *distributed* agents.[5] Resulting plans will usually be incremental, i.e., mention a certain number of next steps to take, but rarely full plans. This is due to the inherent partial observability and disturbances that have to be expected.

In the following, we describe the required components in more detail. It needs a *world model* to represent knowledge about the world and distinguish situations, *task determination* to select the actions to execute, and a *task execution* component which executes and monitors the selected actions. In the given multi-robot context, it also needs to *coordinate* with the other robots to avoid resource conflicts and support cooperation.

Domain Model. The *domain model* encodes everything known a priori about the environment (e.g., expected objects, places) and agents within (e.g., possible tasks or actions). A domain model is particularly relevant for knowledge-based and planning systems, on which we will focus in this chapter.

World Model. To reason about the current situation, to plan into the future what steps to take to achieve a certain goal, and to determine what to do next, the current knowledge about the environment must be represented in a *world model* (WM).

[4]While a global plan is the natural outcome for a centralized planner, a number of local plans is still a possibility, e.g., a number of concurrent planners which use fast communication for asserting plan coherence.

[5]A distributed task-level executive may still produce a global plan, e.g., employing an approach like plan merging [1, 31].

While the domain model encodes off-line information known before the tasks start, the WM represents the knowledge or beliefs about the actual current state. In the anticipated industrial scenarios it typically represents only a partial view of the environment.

Task Determination. With the world model as basis the task-level executive can distinguish different situations to choose a suitable task. Thus certain situations have to be encoded in terms of the knowledge in the world model that determine what goals to pursue, which tasks to achieve, and which actions to take.

There is a plethora of methods for this action selection, which we will describe in more detail in Sects. 3 and 5.

Multi-Robot Coordination. Regardless of which concept is used in the task determination, it is important to consider parallel execution by multiple mobile robots as it allows to speed up and scale the production process by increasing the number of used robots. Furthermore a group of robots is more flexible and robust than a single one because it can, for example, compensate for a robot needing maintenance and thus being unavailable.

When considering a group of robots, their task determination and execution has to be coordinated. Some method of synchronization is required to avoid conflicts. First, robots need to communicate their current intentions to avoid redundant work of multiple robots trying to achieve the same goal. Second, robots need to allocate resources like processing stations for exclusive use by a certain robot during a specified time interval.

Task Execution. Once the robot has decided for a certain task to perform, it needs to execute the involved actions. Here it is advisable to make use of a layered architecture that separates the task-level executive from the low-level system by a mid-level reactive behavior layer. This layer allows to encapsulate many details like parameter estimation that would increase the burden on the upper layer considerably.

An example layered architecture is shown in Fig. 1 (as used in Fawkes [51]). At the top the task-level executive constitutes the reasoning layer. It calls reactive skills encoded as state machines in the mid-level layer, which makes use of low-level functional components such as self-localization and motion. The different components communicate between and inside layers by writing on and reading from a blackboard or by using ROS topics. During the execution of steps, it is necessary to perform monitoring to react to unexpected changes and failures.

Fault Tolerance. To preserve the autonomy of a CPS in case of faults, which are usually impossible to preclude completely, it is important to have fault tolerance in the task-level executive. Failures that can arise are an incorrect world model, e.g., due to uncertainty in sensor data or wrong perception results, or unrecognized problems in the physical actions of the robot, e.g., a robot dropping a work-piece while driving. The fault has to be recognized as soon as possible to prevent a tail of counterproductive decisions based on wrong knowledge; early recognition

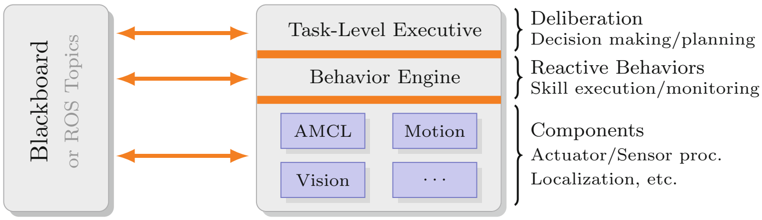


Fig. 1 Behavior layer separation and blackboard communication used in Fawkes

simplifies the identification of what went wrong. Fault tolerance also means to be able to compensate for robots undergoing maintenance or being removed from the fleet.

3 Achieving CPS Autonomy in Smart Factories

In this section, we explore possible solutions and the current state of the art for components required for a task-level executive that provides autonomy to CPS in a smart factory. This is not an exhaustive overview, but draws from the vast pool of literature and methods in particular regarding knowledge-based systems.

Domain and World Model. The domain model is typically very specific to the approach chosen for task determination and execution as the way the knowledge is represented must be supported by these components. For example, for planning approaches a common specification approach is the Planning Domain Definition Language (PDDL [16, 18, 43]). It supports several degrees of fidelity, from simple STRIPS-like operators [14] to ADL [54] and temporal specifications [16]. Besides systems that take PDDL as input, the input languages are often specific to the reasoning system used.

The world model (WM) represents the current belief about the environment. For the world model, the situation can have more facets. First, an *intrinsic world model* (iWM) is part of the task-level executive, for example, in the form of an internal fact or knowledge base. This is because the reasoning and planning algorithms operate on specific data structures in which the known information must be kept. Second, an *external world model* (eWM) implements the world model as a separate component. It can offer several advantages like own reasoning and classification services. For example, in KnowRob [64], ontologies can provide information about where to expect certain types of objects. Mason et al. [41] use a semantic eWM to detect changes in the environment over time. Furthermore, an eWM can allow for more general exchange and synchronization in a multi-robot context, independent of the actual task determination mechanism (e.g., [59]). Third, hybrid approaches, where an iWM is used combined with external synchronization strategies are also possible. The drawbacks of an eWM are the communication overhead between the reasoning component and the WM, which also sometimes means replicating parts of the eWM in another iWM.

Task-Level Executive: Task Determination and Execution. Typical approaches for task-level executives can be roughly divided in three categories [49]: state machine-based controllers like SMACH [2] or XABSL [39], reasoning systems from Procedural Reasoning Systems [28] or rule-based agents like CLIPS [46] to more formal approaches like [38], and finally automated planning systems [20] with varying complexity and modeling requirements [17], for example based on PDDL [43] and its various extensions. There are also hybrid systems integrating aspects of more categories like integrating PDDL-based planning into [7, 27]. The chosen approach depends on the modeling and expressiveness requirements of the application domain, but also on the experience and background of the implementer.

Furthermore, in a multi-robot setting the problem arises to efficiently and effectively distribute the overall workload to the fleet of robots. This can be done by dividing the problem into task decomposition and task allocation [3]. Task decomposition refers to how the overall mission task of the robot team can be decomposed into single sub-tasks that can be independently completed. Task allocation describes the process in which the decomposed sub-tasks are coordinated among individual robots [36]. Both, task decomposition and task allocation can be accomplished in a centralized or distributed manner. Distributed approaches are robust to crashes or temporary malfunctions by increased system redundancy. By distributing task allocation among multiple robots, *distributed* approaches can cope with malfunctions of individual robots as the remaining robots continue operation.

In contrast to distributed coordination approaches, the advantage of *centralized approaches* lies in potentially more efficient plans, since the interaction of robots can be taken into account. Problematic is that all information must be communicated to the central station, this in particular often means more detailed communication than in the distributed case, where only (partial) world model updates must be sent, e.g., to announce that a task is being performed or which workpiece is at a given station. In order to consider the overall fleet, a potentially more complex planning problem must be solved, as more actors and possible combinations must be taken into account for deeper time horizons, i.e., to actually recognize the potential for cooperation, plans must consider multiple steps into the future for multiple actors, leading to a high branching factor and thus planning complexity. The advantage of tightly integrated plans comes at the expense of system reliability as the overall system is dependent on the operation of one central task planning unit. To overcome this shortcoming, approaches exist that can alter the coordination task to different robots. This is what Iocchi et al. [29] call *weak centralization*. In contrast to *strong centralization* (with one fix coordination unit), weak centralization allows for passing the coordination task among different robots. Thus, the multi-robot system can maintain operation in case of a crash of the coordinating robot as another robot takes over the coordination task.

Within distributed approaches, service-oriented architectures are among the most popular in production contexts. One of the first works that introduces the paradigm of service orientation was done by Jammes and Smit in the Sirena project [30]. In service-oriented architectures individual production resources offer a functionality as a service that can fulfill a certain task. These services then interconnect and form services of higher functionalities. An approach of service interconnection are economic approaches such as market- or auction-based approaches. A so called *free market architecture* for coordination of a team of robots is introduced by Stentz and Dias [62]. Within this scenario, the overall task is centrally decomposed into sub-tasks for which revenue and cost functions are defined. The robots then place bids for tasks and negotiate to get tasks awarded. Auction-based approaches to multi-robot coordination are discussed by Gerkey and Mataric [19]. The authors show that distributed negotiation mechanisms within first-price auctions can be effective for task allocation in multi-robot teams. The difference between a market-based approach and the auction-based approach is that auction-based approaches use bids that are based on the individual costs of the robots where market-based approaches take into account both, the costs as well as the benefits of the robots. Kamagaew et al. developed a fully distributed coordination method for multi-robot teams in logistics [33]. The approach is called *cellular logistics* as it was inspired by biological swarm concepts. In the cellular logistic scenario task planning is split into three layers on each individual robot. The lowest layer is responsible for short-term tasks that are coupled to the hardware of the robot. An operational layer is responsible for mid-term tasks and guarantees robustness of task execution. The highest layer is responsible for long-term operation and task planning and realizes the swarm intelligence of the system. If a new transportation task is announced, an auction-based approach is used for task allocation. Bids correspond to path costs transportation sequence as well as the internal state of the robot.

A well known example for strong centralized multi-robot coordination is the KIVA system [67] which is mainly developed for warehouse logistics, but can also be applied to production logistic tasks. In the KIVA system, a central job planner receives customer orders and allocates tasks to robots using utility-based heuristics, where the utility corresponds to the cost for the warehouse owner. Based on these estimates, tasks are allocated to individual robots. Task planning is then done by the individual robots. As the solution space can grow large with the number of robots fully centralized planning can become infeasible. Centralized task allocation by the job planner and distributed task planning on the robots can make the overall problem tractable.

The required *task execution* capabilities depend on the output of the task determination. For (single-step) incremental approaches, it could mean executing single actions. This is typically closely interwoven with the determination component in the task-level executive, an example is a CLIPS-based agent [46]. Full or incremental plans must be executed step by step, employing execution monitoring to verify that the remaining plan is still feasible, and otherwise re-plan. An example for this strategy is Continual Planning [6, 27].

Multi-Robot Coordination. To coordinate resource usage and task assignments, a mechanism for coordination is required. In a centralized approach, a central host may coordinate the use of resources, e.g., by providing an access token only to a single robot at a time.

A common approach of weak centralization is temporal leader election. Leader election algorithms are especially popular in the field of formation forming and exploration where leader-follower paradigms are frequently used for coordination. A very simple leader-election is used in [10] as well as [66] for temporal coordination of multiple robots. Beside its application in weak centralized approaches, the concept of leader-election can also be used for distributed task coordination to multi robots. For example Guerrero and Oliver use a market-based approach with temporal leader election: In case a robot discovers a task, the robot becomes leader of the task, decomposes the task and draws a plan on how many robots are needed and how the sub-tasks are allocated to the robots. The leader then holds an auction in which they search for other robots that take over sub-tasks [24]. Thus, a temporal coalition is formed in which the tasks are allocated to robots.

An approach that directly combines task determination and coordination is plan merging [1, 31]. Robots incrementally extend a global plan in a way that it remains consistent at any one time. Whenever a robot plans, it advertises its plan to the other robots and collects plans from the other robots that affect shared resources. The robot then *merges* its plan with the plans of the other robots without modifying the other robots plans. In larger scenarios, the robot only needs to consider a local sub-set of the global plan, i.e., resources and robots in its vicinity. By this, the global plan is not necessarily explicitly represented in full.

Fault Tolerance. To account for unexpected events and to react to disturbances, methods are required to detect, analyze, and handle a fault.

This component may be an execution monitoring component that explicitly analyzes and verifies the execution of actions and tasks. There are several different ways, how this can be implemented which can roughly be classified to one or more of three approaches, namely analytical, data-driven, and knowledge-based ones [55]. In analytical models, mathematical models (often constructed from first principals like physics) are used to describe the expected behavior and look for deviations. Data-driven approaches often build on statistical methods to learn expected outcomes and spot unexpected values. Finally, knowledge-based systems capture the knowledge of human experts and encode it in a way that it can be analyzed automatically.

There exist also approaches for detecting general system faults, i.e., faults that can be intrinsic to the system without direct outside world influences or independent from action execution. For such faults, also a variety of systems exists [65].

It is uncommon to see intrinsic fault detection systems in robot systems. While they would definitely be useful because the overhead incurred by modeling the system behavior is high. Execution monitoring, however, is a typical requirement as most failures of a robot occur while executing actions.

4 The RoboCup Logistics League (RCLL)

RoboCup [34] is an international initiative to foster research in the field of robotics and artificial intelligence. Besides robotic soccer, RoboCup also features application-oriented leagues which serve as common testbeds to compare research results. Among these, the industry-oriented RoboCup Logistics League[6] (RCLL) tackles the problem of production logistics in a smart factory. Groups of three robots have to plan, execute, and optimize the material flow and deliver products according to dynamic orders in a simplified factory. The challenge consists of creating and adjusting a production plan and coordinate the group [45].

The RCLL competition takes place on a field of 12 m × 6 m partially surrounded by walls (Fig. 2). Two teams of up to three robots each are playing at the same time. The game is controlled by the *referee box* (*refbox*) [47], a central software component for sending orders to the robots, controlling the machines and collecting teams' points. Additionally, log messages and game reports are sent to the refbox, which allows for detailed game analysis and benchmarks. It is also used in a simulation of the RCLL [69]. After the game is started, no manual interference is allowed, robots receive instructions only from the refbox and must act fully autonomously. The robots must plan and coordinate their actions to efficiently fulfill their mission (cf. [49] for a characterization of the RCLL as a planning domain). The robots communicate among each other and with the refbox through wifi. Communication delays and interruptions are common and must be handled gracefully.

Each team has an exclusive set of six machines of four different types of Modular Production System (MPS) stations. The refbox randomly assigns a zone of 2 m × 1.5 m to each station (position and orientation also are random within the zone). Each station accepts input on one side and provides processed workpieces on the opposite side. Machines are equipped with markers that uniquely identify the station and side. A signal light indicates the current status, such as "ready," "producing," and "out-of-order."

Fig. 2 Teams carologistics (robots with additional laptop) and solidus (pink parts) during the RCLL finals at RoboCup 2015

[6] RoboCup Logistics League website: http://www.robocup-logistics.org.

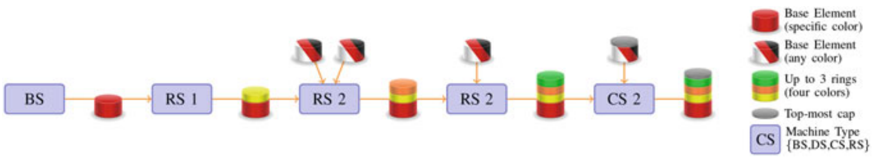


Fig. 3 Steps for the production of a high complexity product in RCLL 2015 (legend on the *right*)

The game is split into two major phases. In the *exploration phase* robots need to roam the environment and discover machines assigned to their teams. For such machines, it must then report the marker ID, position zone, and light signal code. For correct reports, robots are awarded points, for incorrect reports negative points are scored.

In the *production phase*, the goal is to fulfill orders according to a randomized schedule by refining raw material through several processing steps and eventually delivering it. An order states the product to be produced, defined as a workpiece consisting of a cup-like cylindrical colored base, zero, or up to three colored rings, and a cap. An exemplary production chain for a product of the highest complexity is shown in Fig. 3. All stations require communication through the refbox to prepare and parametrize them for the following production step. Using several types of machines, the robots have to assemble specified final products.

The robot used in the competition is the Robotino 3 by Festo (see Fig. 4 for a Robotino 3 in front of an MPS). It features a round base with a holonomic drive, infrared short-distance sensors, and a webcam. Teams are allowed to add additional sensors and computing devices as well as a gripper for workpiece handling.

In 2016, the *RoboCup Industrial* umbrella league [47] was founded which is intended to serve as a coordinating body for industry-inspired RoboCup competitions.

Fig. 4 Robot approaching a ring station in production (Color figure online)

In this context, the RCLL focuses on multi-robot reasoning, planning, and scheduling. Another sub-league is @Work [26], which focuses on mobile manipulation and perception in a manufacturing context.

5 Case Studies on Task-Level Executives for the RCLL

Given our general considerations, what methods may be used to achieve autonomy (Sect. 3), we describe specific case studies of implemented task-level executives for the RCLL and give an outlook on possible future topics. The approaches have all been implemented and evaluated by the authors. They are presented in order of sophistication. First, a CLIPS-based incremental reasoning agent, which has been developed for four years by a number of individuals, has the most extensive feature set. Two strategies have been developed using OpenPRS to the stage that they were capable of playing a very large number of automated games with each other and the CLIPS-based agent. Finally, we describe a prototypical domain specification using the agent programming language and run-time system YAGI.

All systems have been implemented based on the Fawkes Robot Software Framework. It provides the necessary functional components and integration with the Gazebo-based simulation [69]. They all use the very same basic actions which are defined and executed through the Lua-Based Behavior Engine [44]. The full software stacks for 2014 and 2015 have been released as open source software[7] [51].

5.1 CLIPS-Based Agent Program

The first and most sophisticated approach [46] is based on the rule-based system CLIPS. The agent has been used successfully for several years in the RCLL. It evaluates available information, decides for actions to take to achieve the desired goal, and executes and monitors such actions. In general, it encodes certain situations that it evaluates whenever a robot is currently idle if any applies, and then decides for the next action to take. It bases this decision on its internal world model that is shared with other robots.

CLIPS [68] is a rule-based production system using forward chaining inference based on the Rete algorithm [15] consisting of three building blocks [21]: a fact base or working memory, the knowledge base, and an inference engine. *Facts* are basic forms representing pieces of information in the fact base. They usually adhere to structured types. The *knowledge base* comprises heuristic knowledge in the form of rules, and procedural knowledge in the form of functions. *Rules* are a core part of

[7]The 2015 stack is available at https://www.fawkesrobotics.org/projects/rcll2015-release/.

the production system. They are composed of an antecedent and consequent. The antecedent is a set of conditions, typically patterns which are a set of restrictions that determine which facts satisfy the condition. If all conditions are satisfied based on the existence, non-existence, or content of facts in the fact base the rule is activated and added to the agenda. The consequent is a series of actions which are executed for the currently selected rule on the agenda, for example to modify the fact base. *Functions* carry procedural knowledge and may have side effects. They can also be implemented in C++. In our framework, we use them to utilize the underlying robot software, for instance to communicate with the reactive behavior layer described below. CLIPS' *inference engine* combines working memory and knowledge base performing fact updates, rule activation, and agenda execution until stability is reached and no more rules are activated. Modifications of the fact base are evaluated if they activate (or deactivate) rules from the knowledge base. Activated rules are put onto the agenda. As there might be multiple active rules at a time, a conflict resolution strategy is required to decide which rule's actions to execute first. In our case, we order rules by their salience, a numeric value where higher value means higher priority. If rules with the same salience are active at a time, they are executed in the order of their activation.

The CLIPS-based agent implements a local, distributed, and incremental multi-robot system. Each agent reasons separately (distributed) about itself (local). This makes the system robust to network outages and allows for graceful degradation always choosing something to do, even if the plan might be less than optimal (when selecting during network downtime). Whenever idle, a robot commits to the next best action it can take (incremental). This makes the approach particularly resilient towards dynamicity (incoming orders) and disturbances (robot or machine breakdowns).

The robot keeps a *world model* (WM) in its fact base. It consists of game information such as active orders and intentions of other robots. The WM is used to perform *task determination*. Tasks are predefined and can only be activated given some conditions encoding specific situations. A task consists of a number of sequential steps. For example, a transport task would involve to fetch a workpiece from one station and then bring it to another machine. Tasks have priorities allowing to resolve conflicts if multiple could be activated. There is no look-ahead beyond the next task at this point. Breaking down the behavior in modular tasks allows the robot group to execute multiple tasks in parallel. This in turn enables cooperation among robots by executing dependent task. This, however, must be encoded explicitly.

Allowing parallel task execution requires *multi-robot coordination* within the team. In the CLIPS-based agent, for such coordination tasks a central master robot is determined dynamically. The leader election concept is inspired by carrier sense multiple access with collision detection (CSMA/CD) [61] which is used, for example, in Ethernet. Before sending, the sender listens if the medium is busy for another participant. If not, it starts sending. If a collision is detected, it waits a short randomized time before re-trying. In terms of our robot system, robots wait for a short randomized time on startup and listen for master announcements ("sensing the medium”). If none is received, an announcement claiming to be the master is sent. If it collides, i.e., if another master announcement is received at the same time, it again waits for a short randomized time and starts over. This rather opportunistic scheme works well with reasonably randomized wait times and for small groups of robots, which holds in the RCLL. The master regularly announces its presence. If this is not detected by a robot, leader election is performed again.

The world model is updated based on observation or interaction with the environment. A specified subset, e.g., workpieces loaded into a machine, is synchronized with the other robots. The master collects and merges the information of all robots, and sends authoritative world model updates. The master is also responsible for ensuring mutual exclusion. Robots can request the acquisition of a lock for a resource like a certain machine or a task (to avoid two robots committing to the same task). The master ensures that there is always only a single lock active at a time.

After the appropriate locks for resources have been acquired the robot can start *executing a task* (or more precisely, its steps. If a step of a task fails, the task is aborted altogether and a new task determined. Executing a step is generic, calling skills from the Behavior Engine. Additional rules may be added to account for specific conditions, e.g., to perform execution monitoring and consider specific failure cases like an order becoming invalid. *Fault tolerance* is achieved by mitigating robot drop outs through the shared world model. If a robot is unreachable for too long, its tasks are marked as incomplete and its locks released. Once a robot has completed maintenance, it can retrieve the world model from its peers. If the robot with the master role drops out, the leader is negotiated again by the remaining robots.

The execution of a skill is shown in Listing 1. Lines 3–7 describe the required conditions, i.e., the robot must have requested the appropriate lock (ll. 3, 7) and ordered the skill execution (ll. 4–6). If the rule is chosen for activation, it updates the current state (ll. 9–10) and the skill state (l. 11) before calling into the low-level system (l. 12). Since locking is central to our coordination strategy, it has been deeply integrated into the skill execution.

```
1  (defrule skill-common-call
2    "Call skill with a set of args."
3    ?s <- (state WAIT-FOR-LOCK)
4    ?ste <- (skill-to-execute (skill ?skill)
5             (args $?args) (state wait-for-lock)
6             (target ?target))
7    (wait-for-lock (res ?target) (state use))
8    =>
9    (retract ?s)
10   (assert (state SKILL-EXECUTION))
11   (modify ?ste (state running))
12   (skill-call ?skill ?args)
13 )
```

Listing 1: CLIPS rule to execute a skill after the robot has required a lock it waited for

Listing 2 depicts an example rule for a local world model change with some changes being synchronized. A skill has finished successfully (ll. 2–3) which belonged to a step of an active task (ll. 4–5). Lines 7–10 bind facts to variables to be modified in the consequent. Lines 13–14 update the world model regarding the currently held object. The synchronized part of the world model is then updated with the new state (ll. 15–17).

The approach has proven to be very efficient (hundreds of thousands of fact base updates are performed during a typical game) and resilient to robot outages. An in-depth evaluation is given in Sect. 6.

5.2 *OpenPRS*

The Procedural Reasoning System (PRS) is a high-level control and supervision framework to represent and execute plans and procedures in dynamic environments [28]. It is based on the belief-desire-intention (BDI) model [5]. A PRS kernel has three main elements: a *database* containing facts representing the belief about the world, a *library of plans* (or procedures) that describe a particular sequence or policy to achieve a certain (sub-)goal, and a *task graph* which is a dynamic set of tasks currently executing.

```
1 (defrule wm-insert-cap-into-cs-final
2   (state SKILL-FINAL)
3   (skill-to-execute (skill bring_product_to) (state final) (target ?mps))
4   (task (name fill-cap))
5   (step (name insert) (state running))
6   ;an inserted puck without cap will be the final product
7   ?mps <- (machine (name ?mps) (loaded-id 0) (produced-id 0))
8   ?cs <- (cap-station (name ?mps))
9   ?hf <- (holding ?product-id)
10  ?pf <- (product (id ?product-id) (cap ?cap))
11  =>
12  (printout t "Inserted base from CS shelf with " ?cap " cap to fill the CS" crlf)
13  (retract ?hf)
14  (assert (holding NONE))
15  (synced-modify ?mps produced-id ?product-id)
16  (synced-modify ?cs cap-loaded ?cap)
17  (synced-modify ?pf cap NONE)
18 )
```

Listing 2: World model change after successfully inserting a workpiece with cap into a cap station to retrieve the cap

As part of a lab course[8] two strategies were implemented using OpenPRS.[9] Students were given the rules of the game, the simulation, a basic OpenPRS integration, and the CLIPS agent as an example. The task was then to design and implement the overall behavior and coordination of the fleet.

---

[8]The lab's website is at https://kbsg.rwth-aachen.de/teaching/WS2014/LabPRoGrAMR.

[9]The OpenPRS project page is available at https://git.openrobots.org/projects/openprs.

The first strategy is a distributed local strategy, and the second a centralized global one. In the local strategy, the overall task of production is decomposed into sub-goals. Each such sub-goal represents a step in production, e.g., transporting a workpiece from one machine to another or delivering a final good. These sub-goals are combined in a global pool, from which the robots draw tasks whenever idle. The program implements a global semaphore among all robots inspired by the mutual exclusion locks of the CLIPS agent. The overall code structure exemplifies a more loose coupling of behavior components in which individual procedures deal with particular sub-goals. The second centralized global strategy has a specific master agent which determines the tasks that need to be accomplished. It then tasks the individual robots with specific jobs, e.g., transportation of a workpiece. The code structure more explicitly resembles an imperative style program with a main loop. OpenPRS provided an effective environment for encoding the RCLL task. However, it lacks some features, for example, the expression "find the machine M_1, for which there is no machine M_2 which is closer to the robot than M_1" cannot be efficiently queried from the fact base, while this is easily possible in CLIPS. It is also not prepared for embedding into existing systems, but rather expects to operate independently and have modules to access other systems. This makes deep and efficient integration much more involved. We refer to Sect. 6 for an evaluation.

Listing 3 shows an action to move to a specific location. The invocation pattern (l. 2) requires that a "goto" was ordered to be achieved (exclamation mark) in the context of the parameter actually referring to a reachable place (l. 3). The action (l. 4) is called through an external function "skill-call."

```
1 (defop goto
2   :invocation (! (goto $place))
3   :context (navgraph-node $place)
4   :action (skill-call "goto" "place" $place)
5   :documentation "Move to given place."
6 )
```

Listing 3: Ops to call a skill to move to a certain place

```
proc exploration()
  while not (exists <$M> in expl_machines) do
    sleep_ms("1000");
  end while
  while phase == {<"EXPLORATION">} do
    if (exists <$M> in expl_machines) then
      pick <$M> from expl_machines such
        explore($M);
      end pick
    end if
  end while
end proc

proc explore($M)
  goto($M);
  read_light();
  exploration_report($M);
  mark_explored($M);
end proc
```

Listing 4: Procedure that implements the exploration main control loop

5.3 YAGI

A more formal approach for high-level control of autonomous robots are languages from the GOLOG [38] family. It has been used, for example, to encode robot behaviors [11]. With its foundations in the Situation Calculus [42, 58], the world evolves from an initial situation due to actions. Properties of the world are represented by fluents, which are situation-dependent predicates. Complex behaviors are described in terms of actions with preconditions (logical formulas that must hold for an action to be executable) and effects. Effects are described by successor state axioms (SSA). They define the evolution of fluents during execution. The precondition and SSAs combined with some domain-independent foundational axioms form the basic action theory (BAT). Many extensions were proposed, for instance to deal with concurrent action execution [9], continuous change [23], or to allow for probabilistic projections [22], or decision-theoretic planning [4] has been used, for example, for robotic soccer or domestic service robots [60].

YAGI[10] [12] is a recent descendant that aims to define a simpler syntax. Typical implementations of come in form of Prolog interpreters.[11] One of the drawbacks is that such implementations heavily build upon Prolog syntax and Prolog and features are often confused. A proof of concept domain specification has been developed for the RCLL [13]. Listing 4 shows sample code for the exploration phase. As long as there are unexplored machines on the playing field and the game is in the exploration phase (ll. 5–6), the procedure explore is invoked with the argument $\$M$ from the not-so-standard “pick” statement. From the set of all possible machines on the field, the “pick” statement chooses one that satisfies the condition of being unexplored so far. The procedure explore instructs the robot to move to that particular machine, decode the light signal, report the machine to the referee box and finally marks the machine as reported. With these few lines of code, it becomes obvious how compactly code can be written with YAGI, at the cost of waiving detailed execution control.

Listing 5 shows the specification of the goto action that makes the robot move to a given place with given precondition and effects. The signal is used to communicate with the underlying back-end integrating YAGI with Fawkes. A complete description of the YAGI interface is given in [13]. While we only have a prototype implementation of the RoboCup agent in YAGI available, in the future, we will develop a complete agents in YAGI.

5.4 Common Behavioral Architecture

The three approaches have been integrated using the same behavioral architecture depicted in Fig. 1. In particular, the middle layer is exactly the same. It is

[10]YAGI stands for “Yet Another Interpreter”, its website is at http://yagi.ist.tugraz.at/.

[11]The canonical interpreter is available at http://www.cs.toronto.edu/cogrobo/main/systems/.

implemented using the Lua-based Behavior Engine [44] which provides the basic behaviors. The lower layer differs, depending on whether the experiments were run on the real robot (CLIPS) or in simulation (CLIPS, OpenPRS, YAGI). This shows a particular strength of the architecture, the ability to focus on the high-level decision making. The simulation has been a crucial testbed for rapid testing and integration enabling to implement such a variety of solutions for the RCLL.

```
action goto($place) external ($status)
precondition:
  <$place> in navgraph_nodes
effect:
  skill_status = {<$status>};
signal:
  "skill-exec-wait goto{place='" + $place + "'}";
end action
```

Listing 5: Action that calls the Behavior Engine to move to a specified place

6 Evaluation

In this section we present evaluation results for our CLIPS-based and OpenPRS-based agent systems. First, we describe an evaluation based on actual championship games that have been analyzed by means of data recorded automatically by the game's referee box. Second, we describe a large-scale automated simulation tournament we performed to have agents repeatedly play each other.

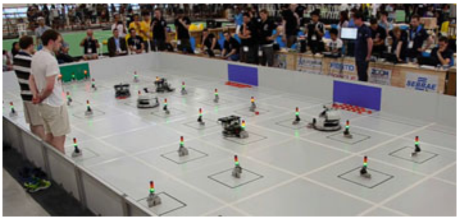
6.1 RoboCup 2014 Evaluation Using the CLIPS-Based Agent

The Carologistics RoboCup team has developed and used the CLIPS-based agent extensively in many competitions [48, 50], winning the German Open and World Championship in 2014 and 2015. Recently, we have presented a data analysis based on 75 GB of refbox data organized using MongoDB of the RoboCup competition 2014 [52]. As an example, we focus on the RoboCup 2014 final of the RCLL between the Carologistics (cyan) and the BBUnits (magenta) which ended with a score of 165–124.[12] While the overall theme of the game was the same in 2014, the game was played on a more constrained field with simpler machines compared to 2015 as shown in Fig. 5. Products were represented by pucks moved on the ground and placed at machines. Each team had 16 machines of 7 different types and there were only three different kinds of products. We will now briefly present the major results, details are given in [52, 53].

Figure 6 shows the machines (M1–M24) grouped per team above the horizontal game-time axis. Each row expresses the machine's state over the course of the

[12]Video of the final is available at https://youtu.be/_iesqH6bNsY.

Fig. 5 Carologistics (Robotino 2 with laptops) and BavarianBendingUnits (Robotino 3) during the RCLL finals at RoboCup 2014



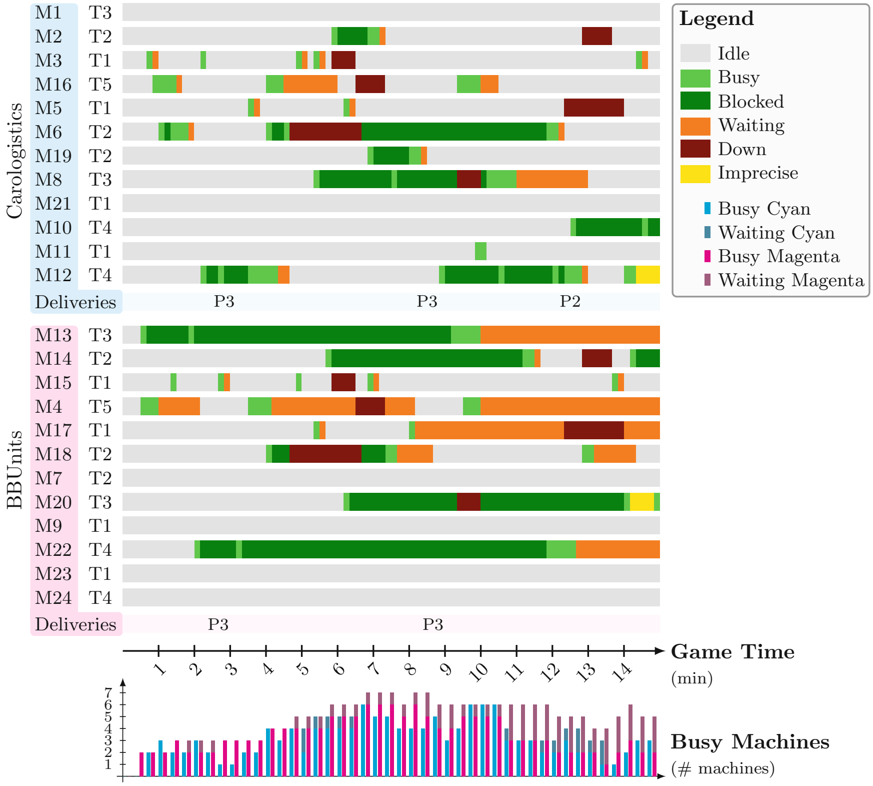


Fig. 6 Machine states over the course of the final game at RoboCup 2014. The *lower graph* shows the occupied machines per 20 s time block (Color figure online)

game. Gray means it is currently idle. Green means that it is actively processing (busy) or blocked while waiting for the next input. After a work order has been completed, the machine waits for the product to be picked up (orange). The machine can be down for maintenance for a limited time (dark red). Sometimes it is used imprecisely (yellow), that is, the product is not placed properly under the

RFID device. The row 'Deliveries' shows products that are delivered at a specific time. Below the time axis, Fig. 4 shows the busy machines over time. Each entry consists of a cyan and magenta column and represents a 20 s period. The height of each column shows the number of machines that are producing (bright team color) or waiting for the product to be retrieved (dark team color).

Figure 6 shows that the production strategies of both teams differ significantly. The cyan team is goal-oriented towards moving workpieces fast through the production steps, indicated by shorter busy and blocked times (green) as well as fewer and very short periods where a machine is waiting for the output to be picked up (orange). The magenta team, on the other hand, has on average more machines occupied (bar chart at the bottom) resulting in longer times that workpieces remain in a machine (dark green areas when waiting for the next input workpiece and orange areas when waiting for output to be picked up).

We recognize that the cyan team achieved a significantly lower throughput time, that is the time that a workpiece requires from the first to the last processing step. The magenta team, however, had a higher number of machines in use at a given time. Only the lower throughput time enabled the cyan team to deliver the complex P2 product in time. Additionally, it allowed for more recycling of material (recyclable material is only available after a more complex machine has completed a work order).

The major observation here is that the grading implicitly favors low throughput over high machine utilizing production—which was not explicitly intended.

Overall, we conclude that the CLIPS-based agent provides a capable framework for modeling medium complex tasks. Especially its ability to closely integrate with C/C++ based underlying software frameworks and comprehensible language, which is used more traditionally in expert systems, allowed to come up with a sophisticated and capable system for this multi-robot task-level executive.

6.2 Automated Simulation Tournament for CLIPS and OpenPRS

The OpenPRS agents were evaluated in a large-scale automated simulation tournament in comparison with the CLIPS-based agent. The simulation is based on the 3D multi-robot simulator Gazebo [35] and models the RCLL [69].

Similar to the CLIPS-based agent evaluation, the OpenPRS evaluation was performed according to the rules of 2014 with simpler machines and fewer product combinations. The simulation is depicted in Fig. 7. The simulation features the same environment agency as the real competition because the machines, whether in simulation or the real world, are controlled by the referee box, an autonomous game controller.[13] For the robots, the simulation exchanges real sensors and actuators by

[13]For details about the referee box we refer to [47] and http://www.robocup-logistics.org/refbox.

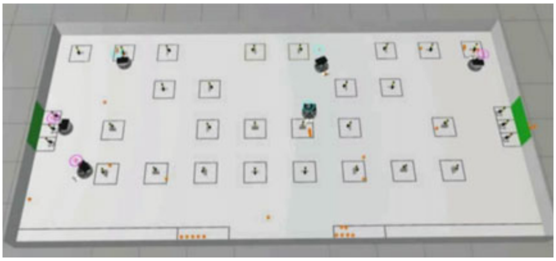


Fig. 7 Simulation tournament used for evaluation with Gazebo and the RCLL in 2014

simulated ones. The high-level programs use the same interface in simulation as on the real robot. Data can be simulated at different levels of abstraction, e.g., the simulation can generate images for running full perception pipelines or simply provide the desired object position. The simulation has been proven as an important tool in the development and evaluation of autonomous CPS in smart factories. During the development, the simulation allows rapid prototyping and testing because of the possibility to test without the physical factory and robots and because of a small setup time. During the evaluation, the simulation allows mass testing to achieve evaluation with high sample sizes and without the need of physical machines and robots. For the mass testing, the simulation is automatically started many times while recording statistics about the simulated games. The games can use different competing teams, configurations and approaches that should be evaluated. Each game uses randomized factory structure and orders by design of the RCLL.

In the simulation tournament shown in Fig. 7, three teams competed in 159 games. The different teams used the CLIPS-based agent and the two described OpenPRS strategies outline above, respectively. All three teams used the same simulated robot, perception and motion components, and basic behaviors. Table 1 shows the summarized results of the simulation tournament by the win amount and achieved points. In addition to the game statistics all simulation runs were recorded for later analysis. The CLIPS agent won the majority of the games against the other teams. This was to be expected given the larger level of sophistication and greater time spent on development. However, the OpenPRS agents were able to perform all games without major faults. The comparison offered some important insights about how to achieve robust and efficient autonomy in the RCLL. Fault tolerance and recovery has been found especially important because some faults could lead to a significantly diminished performance or even a production stop. Here, the fully automatic execution of the simulation runs enforces even more autonomy than the real RCLL because malfunctioning robots cannot be taken out for maintenance. Furthermore, a flexible and fine-grained task allocation leads to faster and more targeted production than a coarse-grained task allocation which assigns a robot to a larger sequence of actions at a time, e.g., accompanying a product through the whole production chain in the factory while waiting at the machines until it is finished. Also the question when to release location-based locks, which ensure two robots do not drive to a position at the same time, has been found important to save

Table 1 Statistics of the mass simulation runs with the amount of games, and point average and standard deviation

	CLIPS	OPRS-1	OPRS-2
Games	104	108	106
Wins	97	13	49
∅ points	120.9	51.0	75.1
σ points	28.2	20.1	23.7

time. Teams which released the lock immediately after the robot finished its job at the location had more collision avoidance trouble than teams waiting to reach a certain minimum distance to the location before releasing it. Another approach that significantly improved the performance was the use of additional storage positions. This allowed improving the utilization of single machines and delivering more products in a short time window compared to let the product wait in the machine output until it is demanded somewhere else. A video highlighting some specifics of the approaches is available at https://youtu.be/5HhOROPLQkY.

7 Conclusion

Introducing cyber-physical systems is a major step in manufacturing industries. To exploit the capabilities they can provide, software components are necessary to provide system intelligence and autonomy, i.e., principled software-driven methods to analyze incoming sensor information, decide on tasks and goals to accomplish and execute the determined actions.

In this chapter, we focused in particular on the decision making and execution components with the primary example of a multi-robot system in smart factory logistics. We have outlined the components necessary to achieve autonomy and presented three case studies with several levels of sophistication and evaluation results that have been implemented to control three autonomous robots in the RoboCup Logistics League (RCLL).

Overall, we conclude that many methods and technologies exist today, which have not been designed with industrial contexts in mind. An extended effort is necessary to bridge the gap between industrial users and researchers from computer science to make these efforts ready for industrial applications. Scenarios like the RCLL can play a role in providing a testbed of comprehensible size that are yet challenging enough for research and development. Our case studies of the RCLL indicate that applying knowledge-based approaches for the task-level executive yield flexible, efficient, and robust systems. The next goal will be to use systems with longer planning horizons to achieve more intrinsic cooperation among the robots.

Acknowledgments We thank the Carologistics RoboCup Team (https://www.carologistics.org) for their tremendous efforts to develop the system with which the case studies could be performed. It would not be without the many students putting endless hours into the project to achieve the seen performance and even implement the various different approaches. In addition to the authors, the team members in 2016 are Mostafa Gomaa, Christoph Henke, Daniel Künster, Nicolas Limpert, Matthias Löbach, Victor Mataré, Tobias Neumann, Johannes Rothe, David Schmidt, and Sebastian Schönitz.

We thank the students participating in the lab course on “Procedural Reasoning on a Group of Adaptive Mobile Robots” (https://kbsg.rwth-aachen.de/teaching/WS2014/LabPRoGrAMR) at the Knowledge-Based Systems Group, RWTH Aachen University, in winter term 2014/2015 for their efforts on implementing the RCLL scenario using OpenPRS. We especially thank Mostafa Gomaa for insightful discussions and preparing the demonstration video (https://youtu.be/5HhOROPLQkY).

T. Niemueller and F. Zwilling were supported by the German National Science Foundation (DFG) research unit *FOR 1513* on Hybrid Reasoning for Intelligent Systems (https://www.hybrid-reasoning.org).

References

1. Alami R, Fleury S, Herrb M, Ingrand F, Robert F (1998) Multi-robot cooperation in the MARTHA project. IEEE Robot Autom Mag 5(1):36–47
2. Bohren J, Cousins S (2010) The SMACH high-level executive. IEEE Robot Autom Mag 17 (4):18–20
3. Botelho SC, Alami R (1999) M+: a scheme for multi-robot cooperation through negotiated task allocation and achievement. In: IEEE international conference on robotics and automation (ICRA)
4. Boutilier C, Reiter R, Soutchanski M, Thrun S (2000) Decision-theoretic, high-level agent programming in the situation calculus. In: Proceedings of AAAI-00. AAAI Press, pp 355–362
5. Bratman ME (1987) Intention, plans, and practical reason. CSLI Publications, Cambridge
6. Brenner M, Nebel B (2009) Continual planning and acting in dynamic multiagent environments. Autonom Agents Multi-Agent Syst 19(3)
7. Claßen J, Röger G, Lakemeyer G, Nebel B (2012) PLATAS—integrating planning and the action language golog. KI-Künstliche Intelligenz 26(1):61–67
8. Clough BT (2002) Metrics, schmetrics! How the heck do you determine a UAV’s autonomy anyway. Technical report, Air Force Research Laboratory, Wright-Patterson AFB
9. De Giacomo G, Lespérance Y, Levesque HJ (2000) ConGolog, a concurrent programming language based on the situation calculus. Artif Intell 121(1–2):109–169
10. Easton K, Burdick J (2005) A coverage algorithm for multi-robot boundary inspection. In: IEEE international conference on robotics and automation (ICRA)
11. Ferrein A, Lakemeyer G (2008) Logic-based robot control in highly dynamic domains. Robot Autonom Syst Spec Issue Semant Knowl Robot 56(11):980–991
12. Ferrein A, Steinbauer G, Vassos S (2012) Action-based imperative programming with YAGI. In: 8th international conference on cognitive robotics
13. Ferrein A, Maier C, Mühlbacher C, Niemueller T, Steinbauer G, Vassos S (2015) Controlling logistics robots with the action-based language yagi. In: IROS workshop on task planning for intelligent robots in service and manufacturing
14. Fikes RE, Nilsson NJ (1972) STRIPS: a new approach to the application of theorem proving to problem solving. Artif Intell 2(3):189–208
15. Forgy CL (1982) Rete: a fast algorithm for the many pattern/many object pattern match problem. Artif Intell 19(1):17–37
16. Fox M, Long D (2003) PDDL2.1: An extension to PDDL for expressing temporal planning domains. J Artif Intell Res 20:61–124
17. Geffner H, Bonet B (2013) A concise introduction to models and methods for automated planning. Synthesis lectures on artificial intelligence and machine learning. Morgan & Claypool, San Rafael

18. Gerevini A, Long D (2005) Plan constraints and preferences in PDDL3. Technical report, Department of Electronics for Automation, University of Brescia, Italy
19. Gerkey BP, Mataric MJ (2002) Sold! Auction methods for multirobot coordination. IEEE Trans Robot Autom 18(5):758–786
20. Ghallab M, Nau D, Traverso P (2004) Automated planning: theory and practice. Elsevier, Amsterdam
21. Giarratano JC (2007) CLIPS reference manuals. http://clipsrules.sf.net/OnlineDocs.html
22. Grosskreutz H (2000) Probabilistic projection and belief update in the pGOLOG framework. In: Second workshop on cognitive robotics (CogRob) at ECAI
23. Grosskreutz H, Lakemeyer G (2001) On-line execution of cc-Golog plans. In: International joint conference on artificial intelligence (IJCAI)
24. Guerrero J, Oliver G (2003) Multi-robot task allocation strategies using auction-like mechanisms. Artif Res Dev Front Artif Intell Appl 100:111–122
25. Hermann M, Pentek T, Otto B (2015) Design principles for industrie 4.0 scenarios: a literature review. Working paper 01/2015, Technische Universität Dortmund
26. Hochgeschwender N, Kraetzschmar GK, Schneider S, Nowak W, Bischoff R, Dwiputra R, Berghofer J, Hegger F (2014) RoboCup@Work: competing for the factory of the future. In: RoboCup symposium
27. Hofmann T, Niemueller T, Claßen J, Lakemeyer G (2016) Continual planning in Golog. In: 30th conference on artificial intelligence (AAAI)
28. Ingrand F, Chatila R, Alami R, Robert F (1996) PRS: a high level supervision and control language for autonomous mobile robots. In: IEEE international conference on robotics and automation (ICRA), vol 1
29. Iocchi L, Nardi D, Salerno M (2001) Reactivity and deliberation: a survey on multi-robot systems. In: Balancing reactivity and social deliberation in multi-agent systems. Springer, Heidelberg
30. Jammes F, Smit H (2005) Service-oriented architectures for devices-the SIRENA view. In: IEEE international conference on industrial informatics (INDIN)
31. Joyeux S, Alami R, Lacroix S, Philippsen R (2009) A plan manager for multi-robot systems. Int J Robot Res 28(2):220–240
32. Kagermann H, Wahlster W, Helbig J (2013) Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Final report, Platform Industrie 4.0
33. Kamagaew A, Stenzel J, Nettsträter A, Ten Hompel M (2011) Concept of cellular transport systems in facility logistics. In: 5th international conference on automation, robotics and applications (ICARA)
34. Kitano H, Asada M, Kuniyoshi Y, Noda I, Osawa E (1997) RoboCup: the robot world cup initiative. In: 1st international conference on autonomous agents
35. Koenig N, Howard A (2004) Design and use paradigms for gazebo, an open-source multi-robot simulator. In: International conference on intelligent robots and systems
36. Korsah GA, Stentz A, Dias MB (2013) A comprehensive taxonomy for multi-robot task allocation. Int J Robot Res 32(12):1495–1512
37. Lee E (2008) Cyber physical systems: design challenges. In: 11th IEEE international symposium on object oriented real-time distributed computing (ISORC)
38. Levesque HJ, Reiter R, Lespérance Y, Lin F, Scherl RB (1997) Golog: a logic programming language for dynamic domains. J Logic Program 31(1–3):59–84
39. Loetzsch M, Risler M, Jungel M (2006) XABSL—a pragmatic approach to behavior engineering. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)
40. Lucke D, Constantinescu C, Westkämper E (2008) Smart factory—a step towards the next generation of manufacturing. In: Manufacturing systems and technologies for the new frontier, The 41st CIRP conference on manufacturing systems
41. Mason J, Marthi B (2012) An object-based semantic world model for long-term change detection and semantic querying. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)
42. McCarthy J (1963) Situations, actions and causal laws. Technical report, Stanford University

43. McDermott D, Ghallab M, Howe A, Knoblock C, Ram A, Veloso M, Weld D, Wilkins D (1998) PDDL—the planning domain definition language. Technical report, AIPS-98 Planning Competition Committee
44. Niemueller T, Ferrein A, Lakemeyer G (2009) A Lua-based behavior engine for controlling the humanoid robot Nao. In: RoboCup symposium
45. Niemueller T, Ewert D, Reuter S, Ferrein A, Jeschke S, Lakemeyer G (2013) RoboCup logistics league sponsored by Festo: a competitive factory automation testbed. In: RoboCup symposium
46. Niemueller T, Lakemeyer G, Ferrein A (2013) Incremental task-level reasoning in a competitive factory automation scenario. In: AAAI spring symposium 2013—designing intelligent robots: reintegrating AI
47. Niemueller T, Lakemeyer G, Ferrein A, Reuter S, Ewert D, Jeschke S, Pensky D, Karras U (2013) Proposal for advancements to the LLSF in 2014 and beyond. In: ICAR—1st workshop on developments in RoboCup leagues
48. Niemueller T, Reuter S, Ewert D, Ferrein A, Jeschke S, Lakemeyer G (2014) Decisive factors for the success of the carologistics RoboCup Team in the RoboCup logistics league 2014. In: RoboCup symposium—champion teams track
49. Niemueller T, Lakemeyer G, Ferrein A (2015) The RoboCup logistics league as a benchmark for planning in robotics. In: WS on planning and robotics (PlanRob) at international conference on Automated planning and scheduling (ICAPS)
50. Niemueller T, Reuter S, Ewert D, Ferrein A, Jeschke S, Lakemeyer G (2015) The carologistics approach to cope with the increased complexity and new challenges of the RoboCup logistics league 2015. In: RoboCup symposium—champion teams track
51. Niemueller T, Reuter S, Ferrein A (2015) Fawkes for the RoboCup logistics league. In: RoboCup symposium 2015—development track
52. Niemueller T, Reuter S, Ferrein A, Jeschke S, Lakemeyer G (2015) Evaluation of the RoboCup logistics league and derived criteria for future competitions. In: RoboCup symposium
53. Niemueller T, Lakemeyer G, Reuter S, Jeschke S, Ferrein A (2017) Benchmarking of cyber-physical systems in industrial robotics—the RoboCup logistics league as a CPS benchmark blueprint. In: Cyber-physical systems—foundations, principles, and applications. Elsevier (to appear)
54. Pednault EP (1989) ADL: exploring the middle ground between STRIPS and the situation calculus. In: Proceedings of the 1st international conference on principles of knowledge representation and reasoning
55. Pettersson O (2005) Execution monitoring in robotics: a survey. Robot Autonom Syst 53 (2):73–88
56. Poovendran R (2010) Cyber-physical systems: close encounters between two parallel worlds [point of view]. Proc IEEE 98(8):1363–1366
57. Radzi NHM, Fox M, Long D (2007) Planning in supply chain optimization problem. In: 26th workshop of the UK planning and scheduling special interest group (PlanSIG-07)
58. Reiter R (2001) Knowledge in action. MIT Press, Cambridge
59. Roth M, Vail D, Veloso M (2003) A real-time world model for multi-robot teams with high-latency communication. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)
60. Schiffer S, Ferrein A, Lakemeyer G (2006) Football is coming home. In: International symposium on practical cognitive agents and robots (PCAR)
61. Shinder D (2001) Computer networking essentials. Cisco Press core series. Cisco Press, Indianapolis
62. Stentz A, Dias MB (1999) A free market architecture for coordinating multiple robots. Technical report, DTIC document
63. Tapas (2016) Tapas—robotics-enabled logistics and assistive services for the transformable factory of the future. http://www.tapas-project.eu. Accessed Mar 14 2016

64. Tenorth M, Beetz M (2013) KnowRob—a knowledge processing infrastructure for cognition-enabled robots. Part 1: the KnowRob system. Int J Robot Res (IJRR) 32:566–590
65. Visinsky M, Cavallaro J, Walker I (1994) Robotic fault detection and fault tolerance: a survey. Reliab Eng Syst Saf 46(2):139–158. http://dx.doi.org/10.1016/0951-8320(94)90132-5, http://www.sciencedirect.com/science/article/pii/0951832094901325
66. Williams K, Burdick J (2006) Multi-robot boundary coverage with plan revision. In: IEEE international conference on robotics and automation (ICRA)
67. Wurman PR, D'Andrea R, Mountz M (2008) Coordinating hundreds of cooperative, autonomous vehicles in warehouses. AI Mag 29(1):9
68. Wygant RM (1989) CLIPS: a powerful development and delivery expert system tool. Comput Ind Eng 17(1–4):546–549
69. Zwilling F, Niemueller T, Lakemeyer G (2014) Simulation for the RoboCup logistics league with Real-World Environment Agency and multi-level abstraction. In: RoboCup symposium

Big Data and Machine Learning for the Smart Factory—Solutions for Condition Monitoring, Diagnosis and Optimization

Alexander Maier, Sebastian Schriegel and Oliver Niggemann

1 Introduction

In this chapter, Machine Learning solutions for Cyber-Physical Systems (CPSs) in a Smart Factory are outlined using productions plants as an example. The increasing complexity of production plants is still a present issue within the industry. Due to an increasing number of product variances, an increasing product complexity and increasing pressure for efficiency in a distributed and globalized production chain, production systems are evolving rapidly: They are becoming modular, can be parameterized and comprise a growing set of sensors. These new features are needed from a production point of view, but they overstrain the capabilities of most users: Too many parameters to optimize and too complex inter-dependencies need to be understood.

According to Lee [8], CPSs "are complex engineering systems that integrate physical, computation and networking, and communication processes. CPSs can be illustrated as physical devices, objects, equipment that is translated into cyberspace as a virtual model. With networking capabilities, the virtual model can monitor and control its physical aspect, while the physical aspect sends data to update its virtual model". The physical world and the cyber world are decoupled by a data-to-feature level, where the data is processed to deliver information in cyberspace at different

A. Maier · S. Schriegel · O. Niggemann (✉)
Fraunhofer IOSB-INA, Langenbruch 6, Lemgo, Germany
e-mail: oliver.niggemann@hs-owl.de

A. Maier
e-mail: alexander.maier@iosb-ina.fraunhofer.de

S. Schriegel
e-mail: sebastian.schriegel@iosb-ina.fraunhofer.de

O. Niggemann
Institute Industrial IT–inIT, Langenbruch 6, Lemgo, Germany

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_18

levels of abstraction, beginning with information about the condition of a machine component leading to the overall throughput and quality risk of a manufacturing line. Therefore, CPSs aim to develop user functions for removing this burden from user capabilities, e.g. for self-diagnosis and self-optimization.

CPSs often tackle the challenge of self-diagnosis and self-optimization in a new and innovative way: Instead of relying on human expertise and additional engineering steps formalizing the necessary knowledge, self-diagnosis and self-optimization is performed in a data-driven way, i.e. by applying ideas from the field of Big Data. Data such as sensor and actuator signals will be stored and analyzed both locally in sensor and control devices, and also globally in Big Data platforms. Based on these data, algorithms will learn models of the normal system behavior. Hence, these models will be used to detect deviations from normal workflow as anomaly (e.g. condition monitoring), find root causes for such anomalies (e.g. diagnosis), calculate optimal parameters and predict the future system behavior (e.g. predictive maintenance, energy optimization).

In this chapter, the following topics are covered:

1. Application of Big Data platforms in CPSs (Sect. 2).
2. Capture relevant data in sufficient quality (Sect. 3).
3. Machine learning algorithms to abstract system observations into such models (Sects. 4 and 5).
4. Usage of models for condition monitoring, predictive maintenance and diagnosis (Sect. 4).
5. Usage of models for the automatic system optimization, e.g. energy consumption (Sect. 5).
6. Application scenarios from the SmartFactoryOWL and from industrial applications (Sect. 6).

2 Big Data in CPSs

The term "Big Data" is not clearly defined in the literature. In common parlance, it describes any data analysis dealing with a huge amount of data. This is not a formal definition, but hit the core of the problem. According to the International Data Corporation (IDC), the volume of data is doubling every two years (http://idcdocserv.com/1678). In year 2014 the world-wide amount of data was about 4.4 Zettabyte. In 2020, the amount will be ten times bigger, around 44 Zettabytes. Werner Vogels, Amazon's chief technology officer, told the BBC (http://www.bbc.com/news/business-26471415): "You can never have too much data—bigger is definitely better. The more data you can collect the finer-grained the results can be." For companies, this means diverse challenges. One of the challenges is the interconnection of the data: A huge amount of the data volumes exists without any connection to other data. Therefore, a challenge to Big Data concepts is to connect data to gain competitive advantages and savings, and to form new business.

Although the best-known applications of Big Data refer to customer data in internet (Google, Facebook, Amazon and others), Big data in manufacturing holds a similar potential. Every used kilowatt hour, from each produced screw to each car, even each switching of a proximity sensor and each change of a temperature sensor generates raw data that holds an enormous potential if it is stored and provided for intelligent analysis. The acquisition, handling and analysis of these data present several challenges.

Due to the heterogeneous data sources providing a different time base, the data has to be synchronized. Several solutions have been proposed for the time synchronization (e.g. IEEE 1588) as well as for the transmission (e.g. Profinet) and semantic annotation of the data (e.g. OPC UA).

The analysis of historical process data throughout the product lifecycle requires new architectures and platforms for dealing with the enormous volume of data, and the variation and speed of data. The data analysis drives the conventional data acquisition and storage to their limits, that is why Big Data platforms are needed.

One of the most used Big Data platforms is the Hadoop ecosystem. A typical Big Data platform is structured as follows: The CPS is connected via a standardized interface (e.g. by OPC UA) with a Hadoop ecosystem. Hadoop itself is a software framework for scalable, distributed computing. The process data is stored in a non-relational database (HBase), which is based on a Hadoop Distributed File System (HDFS). In addition to HBase, a time series database OpenTSDB serves as an interface for data analysis. This database provides simple statistical functions such as mean values, sums or differences that are usually not available in a non-relational data storage. The interfaces of OpenTSDB or Hadoop thus enable data analysis directly on the storage system. Because the algorithms can process the data locally, the volume of a historical dataset does not need to be loaded into a single computer system. Via a web interface, both the data and the calculated results can be visualized (e.g. using Grafana).

Figure 1 illustrates the architecture of a Big Data platform using a Hadoop ecosystem.

While classical condition monitoring is mainly about monitoring of specified signals and checking the crossing of some predefined thresholds to find anomalies, Big data is about discovering the data and finding relationships in the data without specifically looking for something special. This requires the use of novel

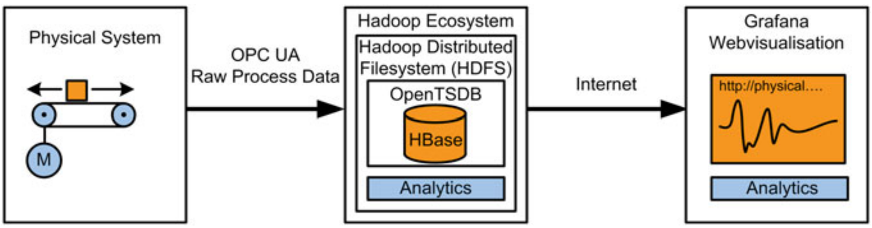


Fig. 1 Example big data architecture

approaches and algorithms. A main issue is the high amount of data which possibly is stored in the cloud and has to be transferred via internet. The limited bandwidth would lead to a delayed analysis. Therefore, the in-memory algorithms are required that perform (a part of) the analysis directly on the data hosting server. Further approaches for the speed-up of the analysis point towards the direction of parallelization. An example mentioned here is the MapReduce technology, which divides the problem into several sub-problems that can be parallelized to gain a computation speed increase, more details will follow in Sect. 4.2.

3 Requirements and Challenges to Data Quality

Smart Factories are built from many assets, consisting of a multitude of different components. The Reference Architecture Model Industry 4.0 (RAMI 4.0, for details see: www.plattform-i40.de) includes the hierarchically ordered assets Product, Field Device, Control Device, Station, Work Centers, Enterprise and the Connected World. All of these physical assets generate or apply data over the application's life cycle. The assets are connected by an Integration and Communication Layer to the functional information processing. The Integration and Communication Layer technologies influence the quality of the information about the physical assets. Low sampling rates, unsynchronized and patchy data do not represent the exact physical state/behavior of the asset, and valuable information is lost. Often, small changes in the time series of signals or energy consumption of specific actors caused by wear of the asset point to a future failure and maintenance requirement. Modern industrial communication systems can fulfill these high performance requirements regarding the data acquisition:

- PROFINET IRT offers sampling rates down to 32,15 μs.
- OPC UA includes semantic data annotations, historical data access and the vertical communication including the Internet and Cloud solutions.
- Time synchronization with PTP—Precision Time Protocol (IEEE 1588) enables factory wide data synchronization with a precision of 100 ns and less [15].

But due to different reasons like device availability, costs, allover system integration and interoperability, these technologies today are often not integrated in one application. Many Machines are communicating with PROFINET, but they are not synchronized to other machines and they do not have historical data access. Often a specific gateway copies the data from the PROFINET system to the OPC UA address space—this works, but due to the performance of the Gateway and due to the performance of the (non-real time capable) OPC UA interface, the data quality is lost: The sampling rate decreases; due to jitter and latencies the synchronization accuracy of the data decreases and often data is lost completely due to bandwidth problems. For this reason, the handling of poor process data is a basic feature of machine data analytic solutions.

On top of these challenges in data quality and data acquisition, it depends on the asset (Product, Field Device, Control Device, Station, Work Centers) what kind of model formalism is useful for further functional data analytics: Discrete manufacturing work centers like straight-line machine sequences can be represented using state machines; continuous processes like rotating machines (generators) need modeling formalisms like clustering.

4 Condition Monitoring and Diagnosis

4.1 Anomaly Detection Using Identified Hybrid Timed Automata

Modern production plants in manufacturing industry are mostly programmed using state machines. The advantage is that the meaning of state machines is easy understandable to humans since they think in a similar way. Therefore, the usage of state machines for anomaly detection in CPSs is obvious. In addition, state machines can be identified automatically from observations. Since production plants are mostly dependent on time, this factor has to be considered as well resulting in a timed automaton as used formalism.

While the identification of untimed stochastic automata is an adequately studied research area [1, 3, 11, 16], the identification of discrete-time behavior in CPSs is a new research area, which arose in the last few years. Various algorithms have been developed to study such timed automata, including RTI + [17] and HyBUTLA [10], which also identifies a hybrid timed automaton. As extension, Markov machines can be used if the probability of the next event depends on a number of previous events, an example is given in Ray [13]. An example for the identification of Bayesian networks is given in de Campos [5].

Most of the aforementioned automaton learning algorithms use the state merging approach for the identification of the resulting automaton. Based on the initially constructed prefix tree acceptor, which comprises all prefixes of recorded production cycles, all pairs of states are checked for compatibility in an iterative manner. If two states are compatible according to a specified criterion, they are merged. After the compatibility of all pairs of states has been computed and the compatible states have been merged, the resulting automaton represents the normal behavior of the considered CPS.

The automaton identification algorithms, which are based on the state merging approach mentioned above, work in an offline manner, i.e. all recordings have to be stored and be available for the state merging procedure. The algorithm OTALA, introduced in Maier [9] is the first algorithm for the identification of timed automata in an online manner. It is based on the assumption that each state in the observed CPS can be represented by a signal vector and each signal vector corresponds to one state in the final automaton. Using OTALA for identification, the observations

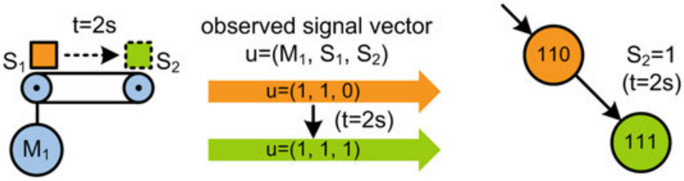


Fig. 2 Online timed automaton learning algorithm (OTALA)

do not have to be stored since each observation is included directly into the final automaton. The basic functionality of the algorithm is shown in Fig. 2.

Basically, the identification of timed automata does not require expert knowledge (except of some parameters which are mostly equal for the different use cases). The states and transitions with the corresponding events are identified by the algorithm. However, structural information about the system (e.g. asynchronous subsystems) is needed to identify one separated model for each asynchronously running subsystem.

The identified automata are finally used for anomaly detection. Much work has been done in this area (c.f. Sarkar et al. [14] for a comprehensive overview). In Vodencarevic et al. [18], the algorithm ANODA (Anomaly Detection Algorithm) has been presented which takes the identified timed automaton as input and detects anomalies in the CPS. The behavior of the CPS can be considered as a path through the identified automaton. Whenever an event is observed in the CPS that cannot be depicted by the automaton in a certain state or if the observed timing does not fit into the identified time range, an anomaly is signaled.

4.2 *Identification of Behavior Models Using Map/Reduce Technology*

Due to the enormous amount of data, conventional data analysis methods are overcharged. In the context of Big Data, extended analysis methods are required to cope with the data flood. As described in the previous sections, the behavior models can be identified from data collected for the system and its components in normal, fault-free operation, using algorithms such as (Hy-) BUTLA or OTALA. These algorithms can be used in Big Data context as well, however, the data handling has to be adapted. Since CPSs create a huge amount of data, the MapReduce technology can be applied to parallelize the identification of behavior models. One possible solution is presented in [19]: OTALA is applied for model learning of the discrete states, and quadratic regression models (QRM) are generated for continuous behavior. Both model learning algorithms have been parallelized applying the MapReduce technology. The MapReduce version of OTALA allows to distribute the workload on $|T|$ nodes, and therefore a speedup is achieved as each transition

T can be processed in parallel to the REDUCE function. For the MapReduce version of QRM, distribution of workload on $|S|$ nodes is possible by processing the states S of the automaton in parallel to the REDUCE function. Furthermore, online algorithms have been proposed, which efficiently handle novel observations to update the models that have been created from large historical data sets.

4.3 Condition Monitoring in Continuous Processes

Condition Monitoring and Anomaly Detection in continuous processes require different approaches than in discrete manufacturing processes. As stated before, Big Data is more than just detecting the crossing of a predefined threshold of a specified signal. It requires unsupervised machine learning techniques, which autonomously find relationships in the data and use these for condition monitoring, for instance.

Clustering approaches are commonly used approaches for these tasks. However, as the data becomes high dimensional, clustering methods (such as DBSCAN, e.g. see Zhu [20] reach their limit since density is more difficult to define in high dimensional space. Therefore, dimension reduction methods are needed to pre-process the data. After reducing the dimensionality of the input data, conventional clustering algorithms can be used to create a behavior model. In [6] self-organizing maps are used to reduce the input dimension and to generate a two dimensional map which visualizes the observed process. A further possibility is the Multidimensional Scaling (MDS, e.g. in Borg and Groenen [2]) which is a set of techniques from the mathematical statistics. The goal is the arrangement of objects and their relation to each other. The farther the objects are from each other, the more dissimilar they are and the closer they are, the more similar they are. There are thus collected information about pairs of objects to identify them to metric information about objects.

The Principal Component Analysis (PCA) is a commonly used method to reduce the dimensionality of the input data. The method basically assumes that features with a low variance provide a small contribution to the final model and therefore can be neglected. To minimize the information loss, the PCA computes the principal components, which are new features that are mostly uncorrelated. The dimensionality of the dataset is then reduced by using the most relevant (see step six in the subsequent procedure for calculating the PCA) of the principal components to describe the dataset. When the PCA is used for visualization purpose, the three first principal components are selected creating a three-dimensional figure. The negligence of the remaining principal components is possible, because most of the variance of the original dataset, i.e. the information, is represented by the first few principal components [7].

The principal components are determined by performing the following six steps:

1. Center and scale the data matrix $\bar{X}$.
2. Compute covariance matrix of data.

3. Calculate the eigenvectors covariance matrix.
4. Sort eigenvectors decreasingly by corresponding eigenvalues.
5. Resulting eigenvector matrix is used as transformation/rotation matrix.
6. Choose x highest dimensions as reduction for PCA.

The PCA is an effective method for dimension reduction. For example, reducing a dataset from 30 signals to 3 principal components (reduction of 90 %), which allows keeping 80 % of the information (corresponds to a loss of 20 %), is a quite effective way of dimensionality reduction.

After the model has been identified using PCA and DBSCAN, it is used for anomaly detection. The anomaly detection is based on the assumption denoted by Chandola et al. [4]: "Normal data instances lie close to their closest cluster centroid, while anomalies are far away from their closest cluster centroid." Therefore, the following two steps are performed for each observation:

1. Transform data from input space to feature space: The transformation matrix (see step 6 from above) is used to transform each incoming observation from the original input space to the reduced feature space.
2. Compute distance to model (see Fig. 3): To determine whether the new observation belongs to the normal operation phase, distances to each point of the model is checked. Several distance functions can be used (e.g. Euclidean or Mahalanobis Distance). The core distance is determined using the Marr Wavelet. In contrast to the Euclidean Distance, the Marr Wavelet crosses the abscissa, the corresponding values are used as core distance.
3. Check thresholds: If the calculated distance exceeds a predefined threshold (identified core distances from the previous step), an anomaly is signaled. As it can be seen in Fig. 3, the first observation was classified as normal, whereas the second observation is identified as an anomaly.

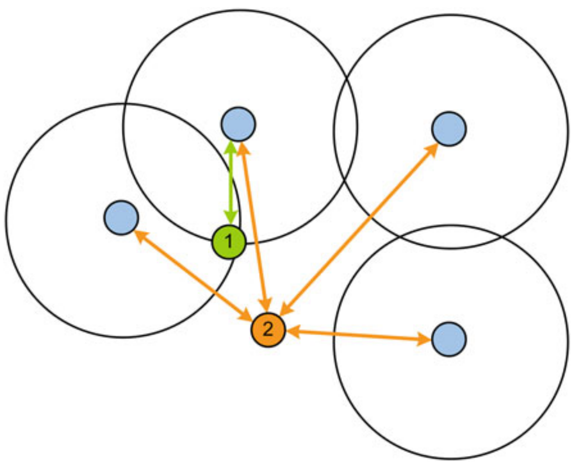


Fig. 3 Distance of observations to normal behavior model

The PCA can also be used for predictive analytics. For this, two core distances have to be used. Crossing the first threshold, a warning is given, crossing the second threshold, an anomaly is signaled.

5 System Optimization

Another application of smart services in manufacturing is the self-optimization of industrial processes. Optimization can be carried out regarding different influencing variables (e.g. time or speed), but in this section we focus on the optimization of the energy consumption. The goal is the analysis and improvement of the performance and efficiency of a manufacturing plant, leading to an optimized operation. Due to increasing energy prices, a special focus for this smart service is the optimization of the energy efficiency in industrial automation systems.

Typically, the optimization of energy efficiency is a manual process, performed by experts of the plant by exchanging old and inefficient drives against new and efficient drives. This is a useful and necessary step, however, it still requires man power and finance investments. Further methods require a manual time planning of the production steps in the Manufacturing Execution System (MES) to obtain an energy-efficient process, or special energy controllers that are typically located at the energy meter and monitor the trend of the energy consumption. If the trend points to unwanted levels, the controller switches off equipment, based on certain priorities and other rules. Typical time periods are in the range of 15–30 min [12].

When it comes to active methods to implement real-time optimization at second or millisecond intervals, only little research has been carried out, because many applications require process parameters to be rapidly adapted to changing operating conditions due to their process dynamics. An active optimization of the process requires models that are able to predict the future process behavior for various parameter combinations in order to determine the optimum process parameters.

Especially in transportation and logistics applications, electrical drive systems are the most voracious energy consumers. Therefore in this application use case, optimization is performed regarding to the energy consumption and load management of electrical drive systems. Figure 4 illustrates an exemplary plant setup of two drives.

The main optimization problem in this application is the minimization of the overall energy consumption of the two converters ($E_{L,WR}$) and motors ($E_{L,mot}$) for time instances $k = 0 \ldots n-1$. Additionally, the energy $E_g(k)$ which is fed back into the internal DC link (difference between regenerative and motive power) has to be minimized. To achieve this, the energy consumption and the energy feedback are weighted with the factor λ according to the primary usage of power feedback (consumption at braking resistances, regenerative feedback to the mains, and temporary storage in capacitors). Altogether, the optimal driving speeds v of the conveying system can be obtained as solution of the following optimization problem:

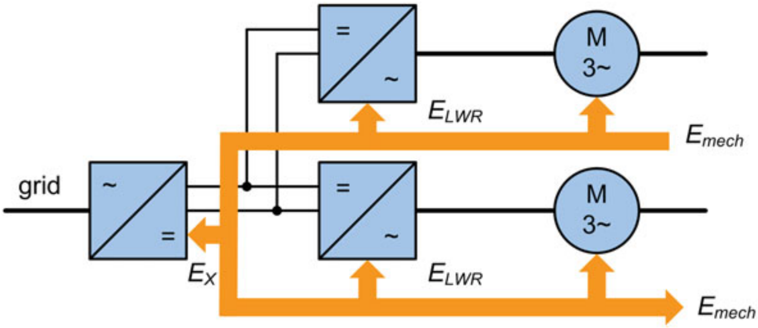


Fig. 4 Two motors connected to a DC circuit

$$v = argmin \left\{ \sum_{k=0}^{n-1} \sum_{c=0}^{n_c-1} y_c(k) + \lambda \sum_{k=0}^{n-1} E_g(k) \right\}$$

Further information regarding the optimization constraints of the energy flows and the sequences of movements (starting position, end position, speed limits, and positioning time) can be found in [19]. The obtained optimization problem is a mixed integer quadratically constrained problem (MIQP). It can be efficiently solved with standard methods. Evaluation results for energy-efficiency optimization of movements in conveying systems verify energy savings of around 7 %. The common optimization of the two drives leads to a motion profile in which the first drive is slowed down just during the acceleration phase of the second drive. In this particular application case, the complete regenerative energy of the first drive has been recovered as input for the second drive. Thus, the automatic coordination of motor speeds prevents undesired energy feedback.

6 Smart Services and Applications

The described technologies, data-driven condition monitoring and optimization of machines or production lines are the basic feature for new technical services (also known as Smart Services), like remote machine operation and new business models. Today, the most usual life cycle of a machine is: The machine builder constructs and sells the machine to the producing company. The producing company operates the machines and produces their products. But the machine-operator is not the overall machine expert (that is the machine builder). Because of this reason, machines are not often driven with the best operation parameters: best product quality, high output, save material and energy, ecology. Summarized: the total production costs are high.

A data-driven condition monitoring and optimization of machines enables two basic features: The machines operate more independent from human operators than before, and the remote operation with machine experts and process experts becomes possible. Due to this, the producing (product) company does not have to be the operator any more. The machine builder could fulfill the role of the machine operator. With this the business model can be transformed: The machine itself is no longer sold to the producing company, it remains the machine builder. Production capacities are leased to the product company. With this, the producing company can concentrate to their main competencies (products) and save investment cost. With this, the machine builder implements a complete new business model, and can use his know-how in their own machine for (1) the construction of new machines and (2, new) for efficient operation of machines.

Fraunhofer evaluated the technical potential of data-driven Smart Services and Applications like condition monitoring and optimization in production lines at various companies like Miele, Wesergold, Audi and Deutsche Windtechnik. These examples are described below. Problem statements for each application are given and the data-driven Smart Service solution is explained.

Miele is a manufacturer of domestic appliances, commercial equipment and fitted kitchens, based in Gütersloh, Germany. Miele is owner-managed company, founded in 1899. The production of dish washers for the very cost-sensitive consumer market has to be as much efficient as production time (machine output), material and energy. Today, process, production and product experts at Miele are running process and machine optimization based on their specific product and machine knowledge. This guarantees maximum product quality, which is the most important objective at Miele. But for an efficient production, the machine optimization has to involve aims like a high machine output, material and energy savings. Often, these are conflicting objectives that cannot be solved by a specific expert. Because of this reason, Fraunhofer evaluated a data-driven Smart Services for Miele. The machine signals of specific production processes were acquired and learned using the machine learning algorithm OTALA. The result was a timed discrete automaton. The automaton showed the time variance for specific process states. Based on this knowledge, better parameters for a higher output were identified.

Tönsmeier is a German company with more than 70 locations, 3000 employees, 1100 vehicles and 30 processing facilities in the waste management industry. The headquarter is located in Porta Westfalica, Germany. Tönsmeier was founded in 1927 by Karl Tönsmeier. In the processing facilities, Tönsmeier is sorting waste with the objective of recycling. The availability of sorting machines is critical because the input of the machine (waste) is often very different and causes failures like plugging and stoppage of the whole sorting machine. With the aim of the prediction and preventing of the plugging problem, Fraunhofer evaluated the performance of data-driven real-time condition monitoring for Tönsmeier.

At the automobile manufacturer Audi, Fraunhofer tested machine learning algorithms with the objective of functionality checks in new production systems. The production systems at Audi are very complex. In the start-up phase of new systems, a lot of time is needed for failure detection and system tests. By optimizing this start-up phase, Fraunhofer evaluated a data-driven assistance system that collects process data from the PROFINET communication of machines, and learns the machine behavior as a discrete state machine. The state machine is a reduced visualization of the complex machine behavior. This simple visualization is used by Audi process engineers searching problems and helps to program the PLCs, and it is also used for acceptance and certification tests of the machines at the end of the start-up phase. This Smart Service is a machine start-up assistance system that helps decreasing start-up phases and increasing the PLC programming quality.

In a research project, Fraunhofer and Resolto implemented a condition monitoring systems that is able to predict ware in wind power plants. The objective was to detect even small changes in processes before they mature into a system error. The system was tested in a wind power plant of the project partner Deutsche Windtechnik. All data of the wind turbines (as sensor values, energy data, weather data, etc.) were collected and analyzed by the self-learning system to model normal behavior of the process. By the help of an adapter and the industrial communication protocol OPC UA, the wind turbines were connected to the Internet. The process data were transmitted to a central server making data analysis. The analysis algorithms were applied to detect discrepancies between the real-time data and the identified model. The models were generated by a principal component analysis.

7 Summary and Outlook

One of the biggest challenges in Big Data is the generation of added value from data obtained in CPS.

In this chapter, we presented some algorithms that address the challenges of Big Data and automatically learn models of normal behavior. The identified models are used for condition monitoring, anomaly detection and predictive analytics. The main benefit of the proposed approaches is that only little expert knowledge is required. The behavior models are largely identified automatically, as well as the parameters needed for anomaly detection and predictive analytics.

A further benefit is the possibility of self-optimization. Based on the identified behavior models and given restrictions, new moving profiles in a storage can be calculated.

Various case studies proved the applicability of the presented machine learning approaches for condition monitoring, anomaly detection and optimization.

References

1. Adenis P, Mukherjee K, Ray A (2011) State splitting and state merging in probabilistic finite state automata. In: American control conference, San Francisco, CA, USA, pp 5145–5150
2. Borg I, Groenen P (2003) Modern multidimensional scaling: theory and applications. J Educ Meas 40:277–280. doi:10.1111/j.1745-3984.2003.tb01108.x
3. Carrasco R, Oncina J (1994) Learning stochastic regular grammars by means of a state merging method. In: Carrasco RC, Oncina J (eds) Grammatical inference and application. Lecture notes in computer science. Springer, Heidelberg, pp 139–152
4. Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Comput Surv 41(3), Article 15
5. de Campos LM (2006) A scoring function for learning bayesian networks based on mutual information and conditional independence tests. J Mach Learn Res 7:2149–2187
6. Frey CW (2008) Diagnosis and monitoring of complex industrial processes based on self-organizing maps and watershed transformations. In: IEEE international conference on computational intelligence for measurement systems and applications
7. Joliffe IT (2002) Principal component analysis. Springer, New York
8. Lee J (2015) Smart factory systems. Informatik Spektrum 28(3):230–235
9. Maier A (2014) Online passive learning of timed automata for cyber-physical production systems. In: The 12th IEEE international conference on industrial informatics (INDIN 2014)
10. Niggemann O, Stein B, Vodencarevic A et al (2012) Learning behavior models for hybrid timed systems. In: Twenty-sixth conference on artificial intelligence (AAAI-12), pp 1083–1090
11. Oncina J, Garcia P (1992) Inferring regular languages in polynomial update time. In: Bunke H (ed) Advances in structural and syntactic pattern recognition. Machine perception and artificial intelligence, vol 5. World Scientific, Singapore
12. Palensky P, Dietrich D (2011) Demand side management: demand response, intelligent energy systems, and smart loads. IEEE Trans Ind Inf 7:381–388
13. Ray (2004) Symbolic dynamic analysis of complex systems for anomaly detection. Sig Process 84(7):1115–1130
14. Sarkar S, Sarkar S, Ray A (2013) Data-enabled health management of complex industrial systems, fault detection: classification. Techniques and role in industrial systems. Nova Science Publishers, New York
15. Schumacher M, Wisniewski L, Jasperneite J et al (2013) Node to node synchronization accuracy requirements of dynamic frame packing. In: 2013 international IEEE symposium on precision clock synchronization for measurement, control and communication (ISPCS 2013)
16. Thollard F, Dupont P, de la Higuera C (2000) Probabilistic DFA inference using Kullback-Leibler divergence and minimality. In: Proceedings of the 17th international conference on machine learning. Morgan Kaufmann, San Francisco, pp 975–982
17. Verwer S (2010) Efficient identification of timed automata: theory and practice. Ph.D. thesis, Delft University of Technology
18. Vodencarevic A, Kleine Büning H, Niggemann O et al (2011) Identifying behavior models for process plants. In: Proceedings of the 16th IEEE international conference on emerging technologies and factory automation (ETFA 2011), pp 937–944
19. Windmann S, Niggemann O, Stichweh H (2015) Energy efficiency optimization by automatic coordination of motor speeds in conveying systems. In: IEEE international conference on industrial technology (ICIT 2015)
20. Zhu L, HeF, Tong Y et al (2015) Fault detection and diagnosis of belt weigher using improved DBSCAN and bayesian regularized neural network. Mechanika 21(1):70–77

Overview of the CPS for Smart Factories Project: Deep Learning, Knowledge Acquisition, Anomaly Detection and Intelligent User Interfaces

Daniel Sonntag, Sonja Zillner, Patrick van der Smagt and András Lörincz

1 Introduction

A tight cooperation of automation and IT vendors should enable sustainable business models supporting the European manufacturing sector to manage its increasingly complex, inter-organizational production networks and align them efficiently with global supply chains. The individual components will be ready as products or as input for product development. Innovation is supported by evaluated business models and concrete examples for customer business cases.

This should be realized by integrating a platform which uniquely combines cross-enterprise event management (anomaly treatment via deep learning, knowledge management via a semantic portal, intelligent user interfaces) with digital product memory technology and smart object virtualization.

In this chapter, we report on our three milestones of the *CPS for smart factories* activity at EIT Digital, funded by the EU.[1] The three milestones can be summarized as

[1]See http://dfki.de/smartfactories.

D. Sonntag (✉)
DFKI, Saarbrücken, Germany
e-mail: sonntag@dfki.de

S. Zillner
Siemens, Munich, Germany
e-mail: zillner@siemens.com

P. van der Smagt
Fortiss, TUM, Munich, Germany
e-mail: smagt@brml.org

A. Lörincz
ELTE, Budapest, Hungary
e-mail: lorincz@inf.elte.hu

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_19

1. CPS Knowledge engineering: understanding the formal requirements of application cases and their formalization;
2. Implementation of software modules and tuning formal models and rules to test scenarios of anomaly detection in physical environments. This includes functional programming with deep learning capacity, and ontology creation/manipulation/extension via a semantic portal infrastructure and intelligent user interfaces [28];
3. Transfer of modules into industrial settings, first evaluations, and business modeling.

2 Technical Infrastructure

The over-arching challenge to address is to combine cyber-physical system (CPS) safety and performance. CPS is to be understood as a network of interacting elements with physical input and output, forming a system of collaborating computational elements controlling physical entities such as Industry 4.0 factories. While addressing safety challenges, the outcomes include models of the behavior of loops with human operators, in particular how to ensure safety. Beyond failures in the robotic system, humans can also make mistakes, and thus a special desired outcome is a model which accounts for humans as producing anomalies by reacting to predictable maintenance tasks and unpredictable events. Technical advancements include, most notably:

1. a GPU-based deep learning machine learning infrastructure for anomaly treatment and data mining;
2. a smart factories knowledge portal infrastructure for an anomaly instance base (knowledge acquisition and management);
3. model-based predictions with anomaly detection followed by workflow management, including real-time verification and, possibly, machine learning fostering earlier anomaly detection;
4. intelligent user interfaces for expert knowledge acquisition, human behaviour input, and human-robot interaction by using, e.g., vision sensors.

These components are built into an architecture and are to be extended with the characterization of a human collaborator who is also *in the loop* and may also exhibit anomalous behavior. Cyber-physical systems are implemented in human environments. The software/hardware outcome package consists of an anomaly management system, including controllers for smart factories. We focus on both open and closed-loop controllers in the robot domain and reporting/maintenance domain in manufacturing.

The work plan for future Industry 4.0 factories comprises business models for Industry 4.0 technology in order to tackle "unmodeled" anomalies that need to be counteracted. That may happen in diverse ways: the first task is the detection and clustering of the anomaly, followed by modeling by means of human expert domain knowledge, and finally, the computer-assisted optimization, including the extension of the ontology or anomaly dictionary and the related (automated) cost-saving workflow management.

2.1 Deep Learning

Based on recent research results, we exploit deep neural networks for the representation and dimensionality reduction of complex sensory data. Neural networks are often understood as the universal approximators studied in the late 1980 s and early 1990 s, having a few layers of hidden sigmoidal units. It has been shown that the training of such shallow networks is NP-complete [17]. Deep networks (or deep belief networks) have the structure of the above-mentioned neural networks but a larger number of hidden layers; typically, this may go beyond 100 [10]. Efficient methods have been introduced recently, see, e.g., [14, 25]. The new generation of deep neural networks can do more than associating outputs to inputs; they can work in reverse and can generate inputs from representations, see [5] for a survey. Such networks are called autoencoders. They have representational and generalization capabilities far beyond that of many others. Generalization capabilities are excellent, but still: the larger and richer the database, the better is the performance for most application cases. The resurrection of neural networks was caused by three important factors:

- the development of deep learning, including the solutions for the vanishing gradient problem by Restricted Boltzmann Machines [5, 14, 25];
- the increase of data set sizes via crowdsourcing;
- the use of graphics processors (GPUs) in computation, leading to processing speed increase of up to two orders of magnitude. This increase allows for serious hyper-parameter searches even in large data sets, eventually leading to better optimization.

It must be stressed that there is still an important dichotomy between neural network and Bayesian machine learning. For a large part, Bayesian analysis does not apply to nonlinear neural networks and a rigorous mathematical analysis of methods or results is not yet within our possibilities. This used to be the case for deep learning, too, but recent developments provide provable bounds for some networks types [2]. Probabilistic neural networks have been developed [3, 4] and are in wide use. Furthermore, variational approximations that exploit the autoencoder concept gained momentum recently [16, 18].

2.2 Knowledge Acquisition

Production controllers and their contextualization demand for a *real-time* semantic layer for, e.g., the assembly lines in automotive factories, the steel production domain, or the energy domain with smart meter and smart grid analytics applications.

In order to comply with the underlying logic of the daily business operations, we rely on a dedicated semantic model supporting longitudinal access across heterogeneous data sources. The semantic model, including the semantic portal implementations and human-in-the-loop knowledge acquisition, will establish the basis for seamless data integration of all production applications.

For the steel production use case, a semantic mediawiki[2] architecture has been implemented. A major goal of this architecture was to combine static facility models to be stored in an RDF triple store and made accessible to the user via semantically enriched MediaWiki pages with dynamically executed business process models in a largely seamless way. In this specific system architecture of the hot rolling mill (steel usecase) there are two sources of information: a digital pen (or pen based interaction on a corresponding smart phone application) and the Object Memory Server (OMS) [13] for accessing an OMS memory that can be stored in simple and cheap RFID labels. This information is then shared with the semantic mediawiki to provide sensor data from the hot rolling mill parts or the production line parts in real-time. The goal of this knowledge portal development is the realization of a situation-specific adaption of production steps, i.e., process parameters, because of certain potential anomaly events during the manufacturing process. The dynamic behavior of smart machines was realized by a knowledge-based decision making component. The component decides on the actions to be taken in the production process based on information about product variants and machine capabilities described in specifically designed ontologies.

As a foundation of semantic product memories for the internet of things and this smart factories project, see [35]. The development of low-cost, compact digital storage, sensor and radio modules allows us to embed digital memories into products to record those anomaly key events. Such computationally enhanced products can perceive and control their environment, analyze their observations, and communicate with other smart objects and human users. The RFID and semantic portal infrastructure supports the interaction with digital product memories [19] and controlled interaction with digital product memories [12] during an assembly process. In addition, user input data is forwarded to the corresponding servers on which additional handwriting and gesture recognition (active user input) or sensor checking procedures (automatic passive sensor input) are executed. These results (e.g., documented anomalies) are stored in the respective XML documents of the semantic MediaWiki (see Fig. 1).

[2]https://semantic-mediawiki.org.

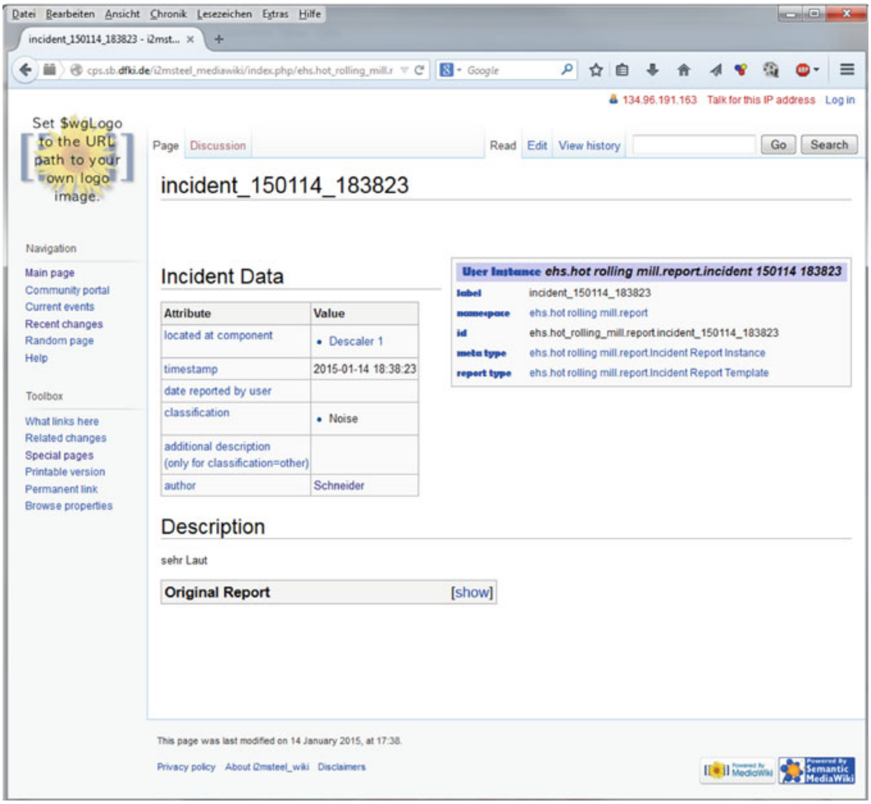


Fig. 1 Example of an automatically generated incident report (anomaly)

2.3 *Anomaly Detection*

Models are required for describing and controlling the dynamic behavior of non-linear plants. Typically, such models are not sufficiently rich, especially if the plant has many degrees of freedom or high-dimensional sensors, or when it is embedded into a complex environment like robotic systems, intelligent vehicles, or any other modern actor-sensor system that we depend on. In such cases, the quality of fault detection deteriorates: too many false positives (i.e., false alarms) make the fault detection useless, while too many false negatives (i.e., unobserved faults) may harm the system. Rather than fully trusting incomplete models, we have put forth a methodology which creates a *probabilistic vector time series model* of the system from the recorded data and detects outliers, also called anomalies with respect to this learned model. This type of detection is notoriously difficult as it is an ill-posed problem. First, the notion of anomaly strongly depends on the domain. Then, the boundary between "normal" and "anomalous" might not be precise and might evolve over time. Also, anomalies might appear normal or be obscured by noise.

Fig. 2 Frame series of the assistant hitting the robot arm

Finally, collecting anomalous data is very difficult, and labeling them is even more so [9].

Two observations are important to make: (i) anomalies are sparse by their very nature, and (ii) in a high-dimensional real-world scenario, it will not be possible to rigorously define "normal" and "anomalous" regions of the data space. We therefore designed an unsupervised approach together with a data collection machinery by using either human interactions (Fig. 2) or machine generated "anomalies" (Fig. 3).

Figure 2: In one step of the data generation process, a probabilistic vector time series model of the system's data is created, and (patterns of) samples that do not fit in the model were conjectured to be the anomalies. We found that searches for anomalies can be made robust by means of Robust Principal Component Analysis [8] if it is combined with group fused Lasso techniques [7] and sparse event filtering [22].

Figure 3: The seven measured joint angle trajectories are plotted in different colors. The arrival of a new desired configuration for the robot arm is visualized by a thin vertical black line. A trajectory segment from the current to the desired configuration is computed and executed. Upon arrival at the desired configuration, a new desired configuration is sampled. In each of the 10 segments there is a probability of 15 % that an anomaly is introduced, visualized by a vertical, light colored bar matching the color of the joint the anomaly was induced on. In this example, the anomaly in joint w1 is clearly visible as an indent in the sequence ranging from about 14 s to 15 s. For an anomaly-free sample, the plot would look the same except that there would be no "command" in the sequence.

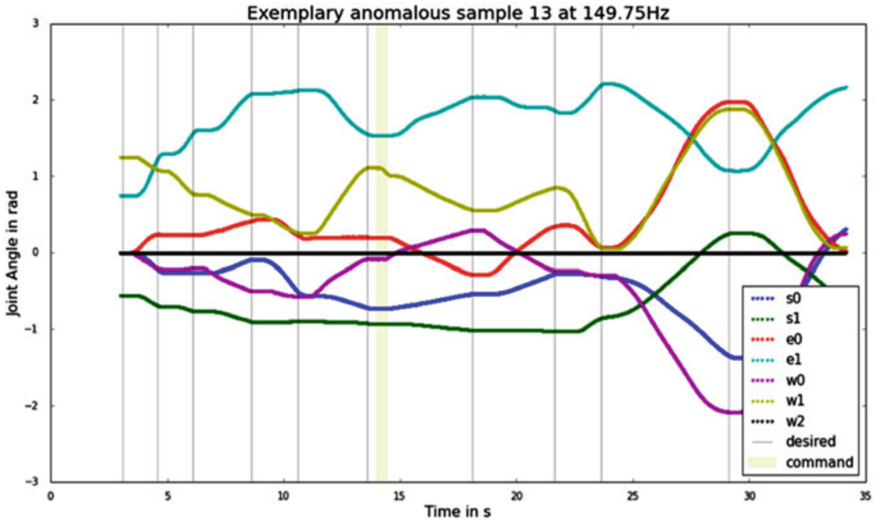


Fig. 3 Exemplary anomalous sample

Although sparse methods are efficient, they are also slow. In order to overcome this critical issue, we developed an architecture that fits the deep learning scheme that we call "Columnar Machine" [20]; this machine can take advantage of the group structure (if present) as in the above-mentioned robust anomaly detection scheme.

Both workflow management and verification require a modular structure and goal-oriented optimization techniques. We developed a theoretical framework for cyber-physical systems with learnable stochastic models of the environment for risk management [32]. The framework meets the constraints of functional programming [21], a desired feature in software development, testing, and verification.

2.4 Intelligent User Interfaces

Internet of Things (IoT) is mainly about connected devices embedded in our everyday environment. Typically, "interaction" in the context of IoT means interfaces which allow people to either monitor or configure IoT devices. Some examples include mobile applications and embedded touchscreens for control of various functions (e.g., lights or control buttons) in environments such as smart factories. In our application cases, humans are an explicit part of the scenario. Traditional graphical interfaces often lead to a clumsy co-existence of human and IoT devices (consider a tablet for remote-controlling a robot arm). Thus, there is a need to investigate what kinds of interaction techniques could provide IoT to be more human-oriented, what role automation and interaction has to play, and how

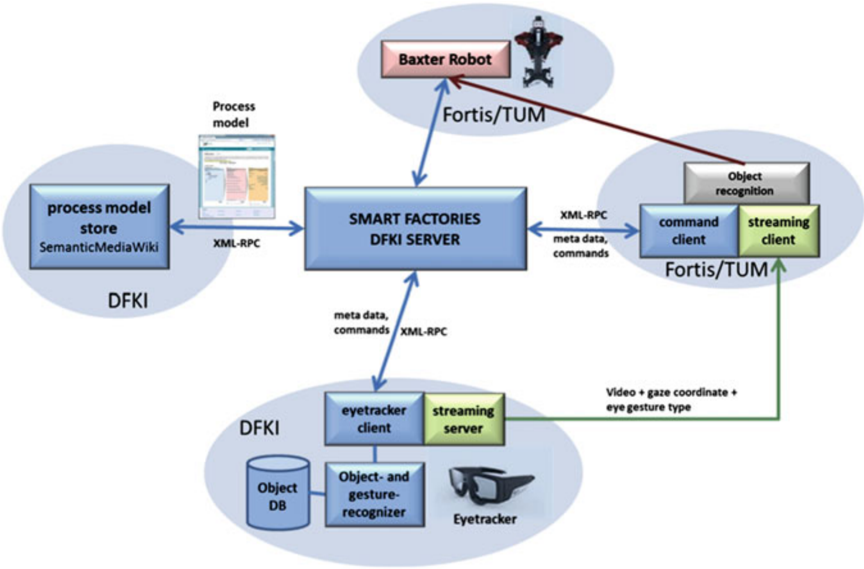


Fig. 4 Intelligent user interface architecture

human-originated data (sensor data for physiological computing) can be used in IoT [33]. Figure 4 shows our intelligent user interface IoT architecture including the semantic mediawiki, the industrial Baxter robot, and the object-detection based anomaly detection.

In many scenarios in modern work practices where robots improve traditional industrial work flows, the co-operation of robots with human workers plays a central role. In this project we focus on direct human-robot-interaction in scenarios where humans and robots interact with the same object (workpiece or tool). Thereby, we address the mutual identification of an object through the human and the robot. It is easy to inform the user about which object the robot is addressing (e.g., by pointing, synthesized speech, explicit action). One of the challenges is to inform the robot about the object the human is attending to without interrupting and affecting the execution of the manual task of the human worker. Figure 5 illustrates the scenario. For the purpose of detecting human intentions automatically, and support activity recognition in general, we developed an eye-tracking system capable of analyzing human eye movements and interpreting specific movement patterns and fixations as eye gestures that were sent to the robot's control system.

To build a base system for experiments leading to such a functionality, it was necessary to develop an operating software for a commercial eye-tracking hardware. Available standard software lacks important features like network transport of gaze points and objects the user gazes at. Also the support of cloud based databases for visual object features that are important in co-operative industrial applications was typically missing in commercial eye-tracking systems. Therefore the

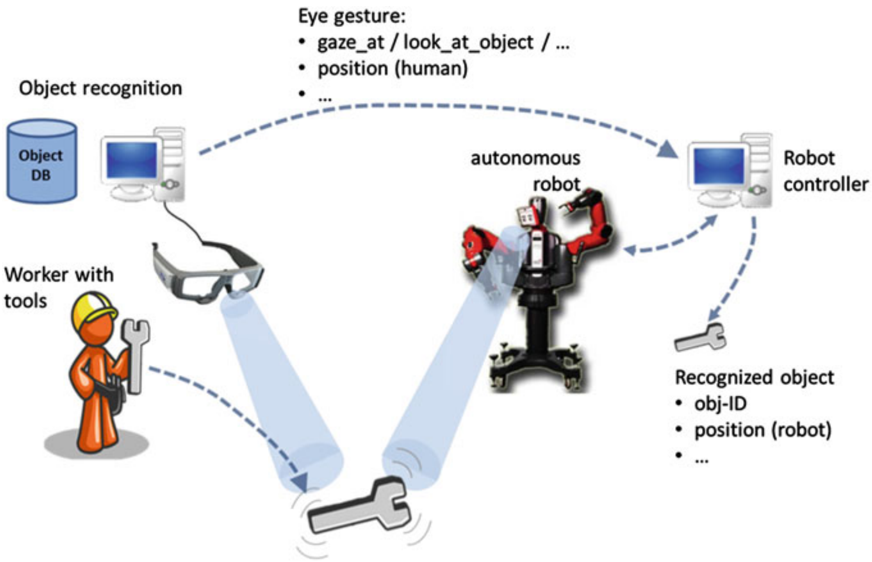


Fig. 5 Overview of the co-operation of robots with human workers

distributed service architecture of the project was extended by components around an object- and eye-gesture-recognition service. This system uses an object database for storing visual feature vectors of given objects, analyzes the user's eye movements, extracts the gaze coordinates, and identifies different eye gestures. The extracted information is then sent to a client via an efficient low level network protocol. The lower part of Fig. 4 visualizes the described components of the system architecture based on [27, 29–31].

In order to allow for vision-based object and activity recognition, we implemented a 3D video annotation tool to provide supervised learning material for the deep learning and anomaly detection infrastructure. Dasiopoulou et al. [11] present an overview of the state-of-the-art in image and video annotation tools. Two new directions are prominent for Industry 4.0 CPS: first, recent work leverages on highly capable devices such as smartphones and tablets that embrace novel interaction paradigms, for example, touch, gesture-based or physical content interaction [26]; we generalize this to multimodal multisensor annotation in the smart factory context. Second, our own previous development LabelMovie, a semi-supervised machine annotation tool with quality assurance and crowd-sourcing options, has been opted for videos (spatio-temporal annotation) [23]. The annotation tool provides a special graphical user interface for multimodal multisensor data and connectors to commercially available sensor systems (e.g., Intel RealSense F200 3D camera, Leap Motion, and Myo). The collected videos are used to learn hand movement for activity recognition in the joint interaction space of a human and a robot.

3 Use Cases

The tight cooperation of automation, cyber-physical systems in production engineering, and IT vendors should enable sustainable business models supporting the European manufacturing and smart grid sector to manage its increasingly complex, inter-organizational production networks and align them efficiently with global supply chains. The individual components of the technical architecture should be made ready as service products or as input for product development. Innovation is supported by evaluated business models and concrete customer business cases, e.g., online anomaly-detection methods in a welding-based production scenario. Recent technological developments include breakthroughs in object memories (smart meter data), big data analysis, and controller software.[3,4]

We include new methods for finding and treating anomalies such as deep neural networks and the semantic portal that can host instance bases of anomalies.

3.1 Industrial Robots and Anomaly Modeling

We exploit online anomaly-detection methods in production scenarios. Quality control in a production scenario is an important attainable goal, and our methodologies carry the potential of obtaining that by combining deep learning with sparse representations [22].

Human-robot interaction requires the detection, interpretation, and prediction of human body movements within context, including hand, arm, head, face, and eye movements that reveal information about the manipulations and thus about the ongoing activity, including the intentions. We have been developing a novel database for hand pose tracking and exploit deep-learning methods for the estimations. The tool is under testing and the data set can be extended if needed. In this project, we use a hand model (libhand, [34]), 3D cameras of different kinds, and a SmartGlove.[5] After initial evaluation it can be said that only the combination of SmartGlove and libhand suits our goals (precise measurements). Data collected with SmartGlove are transferred to libhand and different 3D views are collected. In current work, we explore robotic motion together with facial expression and human body distance (which is required for safety reasons). A facial expression estimation tool is also available to us, based on [15]. Head pose estimation is very precise, the gaze estimation tool has only about 2° of uncertainty, but requires a high-resolution input video.

In many scenarios in modern work practices where robots improve traditional industrial work flows, the co-operation of robots with human workers plays a

[3]http://sites.tcs.com/big-data-study/return-on-investment-in-big-data/.

[4]http://www.greentechmedia.com/research/report/the-soft-grid-2013.

[5]http://www.neofect.com/en/smartglove.

central role. For human-robot interaction, we developed a Unity 3D serious game in order to model realistic scenarios. This game examines human behavior in tasks of divided attention. The game was implemented for the 3D virtual reality headset Oculus Rift, equipped with hand pose estimation 3D camera and the SMI tool for gaze direction estimation. Serious games scripts can be run. The game of the use case was carefully chosen out of many dozens of other games, the main point being to develop a user model, where WCET (worst case execution time) distribution can be estimated. In this game, the need for workflow management becomes straightforward, since attended regions and possibly intentions can be estimated from gaze, together with the registration of the simple manipulation tasks. This is a model for a controller room, where divided attention can be measured under various conditions, including adjustable stress levels. In turn, these games provide a high quality model for CPS with human-in-the-loop problems. A further goal beyond the quantification of the behavior of the human participant is to find methods for proper robotic help in case of anomalies. A number of experiments have been conducted with 10 people and in about 10 sessions. Estimation of WCET is in progress. As a result, this game goes beyond the problem of human-robot collaboration.

3.2 Anomaly Treatment in the Steel Domain

The main focus of this use case scenario is to enable a seamless integration of production and maintenance processes in the context of anomaly treatment. Proper maintenance of industrial plants is of high relevance. It helps to significantly reduce operating costs as well as to improve productivity of the plant operations and the quality of the product. The overall objective of a plant maintenance management system is to ensure the reliability of a plant (component) to perform its functions. Thus, maintenance is seen as any activity that is carried out on a plant or respectively component of a plant in order to ensure that this plant or component of a plant continues to perform its intended function.

However, as of today, the integration of production and maintenance processes and know-how is only addressed and realized in a very limited way. In the past, when the machines have not been that automated, complex and connected, employees from the production site included maintenance task into their daily routines. Only in situations when the handling of the identified failure exceeded their own expertise, external maintenance supports were requested.

Today, with the increase in automatization and digitalization of plants, more and more monitoring and maintenance applications are available. Those monitoring applications are primarily designed to track single and isolated components or parts. Other frequently used maintenance routines are predictive maintenance application servicing single parts or components of plants, such as condition-based monitoring applications for individual plant operations or plant components. In general, those techniques are complemented with preventive maintenance strategies. For example, on a predetermined periodical basis, components of the plant are taken off-line in order to inspect them. Based on the inspection result, repairs are made and the components are put back into operations or the affected components are replaced. Thus, due to the sheer complexity of the underlying processes and operations, this leads to the situation that the particular employees having the highest experience with handling the machines and plant components are no longer actively involved in the maintenance process. In sum, although various maintenance applications and efforts are accomplished on different levels, several shortcomings can be observed:

1. many different monitoring applications, such as predictive maintenance applications, provide important insights about plant components, but do not produce insights covering the comprehensive perspective on the plant performance. The maintenance is focused on local aspects but ignores the plant performance as a whole;
2. the knowledge and expertise of production employees most experienced in handling the plant components are no longer actively involved in the maintenance process;
3. the semantic knowledge about the structure and the basic principles of the plant is not incorporated into the maintenance processes. In particular, the semantics about how plant components are operating and how they are connected with each other as input for the interpretation of local maintenance observations in a global scale is neither available nor used.

In order to overcome the described shortcomings, the general idea of the extended business use case is to seamlessly align human-generated expert know-how with machine-generated maintenance know-how in a semantically consistent manner in order to significantly improve the analytics-based maintenance applications (Fig. 6). The main contribution of the described technical infrastructure for selected extended business scenario is

1. to seamlessly align the local perspective of the large number of monitoring applications, such as predictive maintenance applications, as well as the historical data of the various preventive maintenance strategies into one global and coherent perspective;
2. to establish means for the seamless integration and processing of expert knowledge in the production field;
3. to incorporate the structural knowledge about the plant and its operations by means of a semantic model covering various levels such that the general representation of all necessary (pre-processed) data sources becomes possible; and
4. to use this integrated data source as input for analytics applications aiming to produce new valuable insights as well as to trigger automatically recommended actions for improved overall plant maintenance processes.

As already indicated above, existing plant monitoring and maintenance application can be classified as follows [6]:

- *Breakdown maintenance applications*; i.e., applications that help to fix (the component of) a plant when it is broken. This maintenance routine is reactive

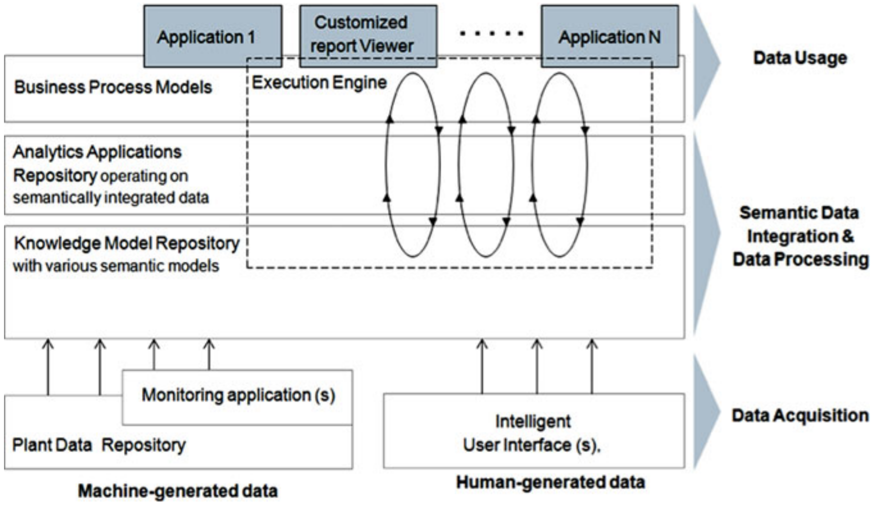


Fig. 6 Overview of the steel domain business scenario

and is only executed when plant equipment needs to be repaired. It is neither based on an underlying routine maintenance task, nor on a scheduled maintenance strategy.

– *Preventive maintenance applications*; i.e., applications that are based on fixed maintenance schedules in order to replace the affected components of a plant.
– *Predictive maintenance application*; i.e., applications for the condition-based monitoring of the plant operations/ components.
– *Proactive maintenance applications*; i.e., application that concentrate on the proactive monitoring of plant operations/ components enhanced by the correction of root causes to component failures.

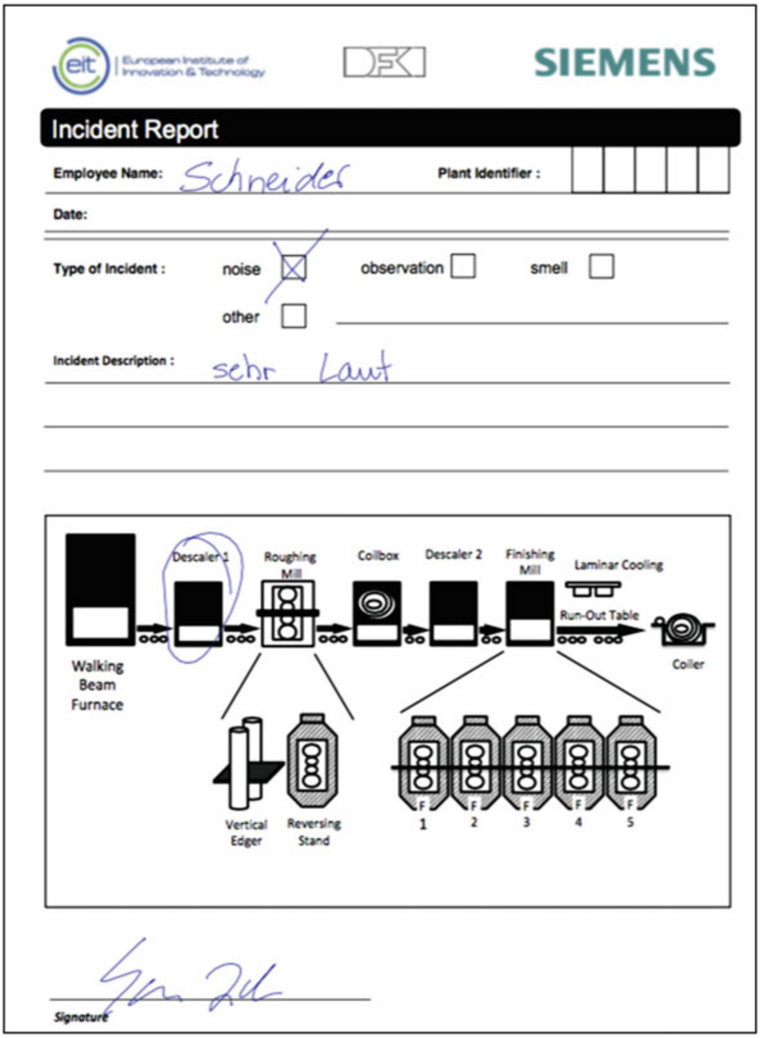
The maintenance application rely on a high degree of automization and digitalization. To the best of our knowledge, there exists no application that allows to seamlessly integrate and process human-based and factory-based monitoring data to improve the overall performance. There are related working papers highlighting the need to align maintenance and production processes. However, besides highlighting the problem scope, those works, e.g., [1], do not provide any concrete solution on how to address these shortcomings. In addition, there exist several research approaches that investigate the use of semantics and ontologies for improving maintenance processes and intelligent fault diagnoses. For instance, [36] introduces an ontology-based reasoning framework for intelligent fault diagnosis of wind turbines. However, this approach neither covers the seamless integration of human-generated data, nor the seamless integration into related business processes. [24] introduces an approach for modeling the semantics of a failure context in order to improve the maintenance support for mobile actors. Although this approach makes use of semantics to formally describe the failure context, the overall applications scenario does not focus on the seamless alignment of human- and machine-generated know-how.

The overall idea of our use case scenario is to establish an application that allows the seamless integration, alignment, processing and analyzing of machine and human-generated monitoring data in order to produce new insights with business value. The machine-generated data originates from the various condition-based monitoring applications or the data repositories of the plan. The human-generated data is captured by semantic-based/ intelligent user interaction applications. In order to realize the extended business scenario, several components are required (Fig. 6).

The *plant data repository* (our Semantic MediaWiki) encompasses all data sources produced and stored in the context of the production process. For instance, the plant data repository collects any historical information about the accomplished production processes (i.e., all accomplished transformations). The data sources can be distributed along the process production chain. A set of *monitoring application(s)* is needed; these continuously measure the condition of component in order to assess whether it will fail during some future period. The data collected in general focus on tracking the condition of particular features, such as vibration analysis, infra-red thermographs, ultrasonic detection, etc. Although the recorded data can be utilized to determine the condition of an isolated component in order to decide about any necessary repairs, the recorded data about possible anomalies will be seamlessly aligned within the adapted incident-report model to establish the basis for aligning human and machine-produced monitoring data for improved plant performance. An *intelligent user interfaces* establishes means to acquire human-produced data in an efficient manner. This semantic data acquisition component, which can be realized as a smart pen application, is aligned with the underlying working routine of the experts in order ensure ease-of use. In addition, the intelligent user interface ensures that all data is captured in semantically annotated form. In general, this is realized by determining the underlying context of the user-input and expressing it with corresponding semantic terminology. In this implementation of the use case scenario, we are using a smart pen application in combination with a customized incident/anomaly report (paper-based). Figure 7 shows the incident report document (hot rolling mill) to be filled out by a maintenance worker in case that an anomaly has been detected.

3.3 Outlook: Anomaly Detection in the Energy Domain

According to related surveys, integrated data analytics in the energy and resources domain promise business impact: in comparison to other industries, companies in the energy and resources industry are expected to generate, amongst others, the highest returns on big data investments. In order to make use of this wealth of data, we have two challenges to tackle:



European Institute of Innovation & Technology

DFKI

SIEMENS

Incident Report

Employee Name: Schneider   Plant Identifier :

Date:

Type of Incident : noise [x] observation [ ] smell [ ]

other [ ]

Incident Description : sehr Laut

Signature

Fig. 7 Incident report document

1. seamless data integration: how to make disparate and incompatible datasets usable, interoperable and valuable across enterprises.
2. data analytics for insights into new products.

By aligning existing data discovery technologies and semantic technologies, new insights in the areas of smart meter and smart grid analytics application will be investigated and prototypically implemented. The focus will be on seamless integration of the Information Technology with the Operational Technology covering all kind of sensors from for example protection devices, via supervisory control and data acquisition (SCADA), etc. The range of analytics applications in the integrated IT/OT world cover, for instance, voltage map generation, outage prevention, optimization applications, fuse dimensioning and asset life, asset performance management, prediction, or fault grid analysis. Potential business applications range from asset performance management, SCADA based data analysis, to outage prevention, etc. Through systematic customer evaluation processes, value propositions will be identified and implemented as prototypical and advanced functionality.

Acknowledgments We thank our project partners in CPS for Smart Factories, funded by EIT Digital (EU, Horizon 2020). See http://dfki.de/smartfactories/.

References

1. der Arbeitsdirektionen Stahl, AEM (2003) Integration von produktion und instandhaltung. Arbeitspapier, Hans-Boeckler-Stiftung
2. Arora S, Bhaskara A, Ge R, Ma T (2013) Provable bounds for learning some deep representations. arXiv preprint arXiv:1310.6343
3. Bayer J, Karl M, Korhammer D, van der Smagt P (2015) Fast adaptive weight noise. arXiv preprint arXiv:1507.05331
4. Bayer J, Osendorfer C, Urban S, van der Smagt P (2013) Training neural networks with implicit variance. In: Lee M, Hirose A, Hou Z-G, Kil RM (eds) Neural information processing. Springer, Heidelberg, pp 132–139
5. Bengio Y (2009) Learning deep architectures for AI. Found Trends® Mach Learn 2(1):1–127
6. Jabar HB, Sdn Bhd, SP (2003) Plant maintenance strategy: key for enhancing profitability (2003)
7. Bleakley K, Vert JP (2011) The group fused lasso for multiple change-point detection. arXiv preprint arXiv:1106.4199
8. Candès EJ, Li X, Ma Y, Wright J (2011) Robust principal component analysis? J ACM (JACM) 58(3):11
9. Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Comput Surv 41(3):15:1–15:58 http://doi.acm.org/10.1145/1541880.1541882
10. Chen T, Goodfellow IJ, Shlens J (2015) Net2net: accelerating learning via knowledge transfer. CoRR abs/1511.05641. http://arxiv.org/abs/1511.05641
11. Dasiopoulou S, Giannakidou E, Litos G, Malasioti P, Kompatsiaris Y (2011) Knowledge-driven multimedia information extraction and ontology evolution. In: Paliouras G, Spyropoulos CD, Tsatsaronis G (eds) A survey of semantic image and video annotation tools. Springer, Heidelberg, pp 196–239
12. Gebhard P (2013) Controlling interaction with digital product memories. In: Wahlster, W. (ed.) SemProM. Cognitive technologies. Springer, Berlin, pp 243–259
13. Haupert J, Schneider M (2013) The object memory server for semantic product memories. In: Wahlster W (ed.) SemProM. Cognitive technologies. Springer, Berlin, pp. 175–189 (2013)
14. Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18:1527–1554
15. Jeni LA, Lorincz A, Szabó Z, Cohn J, Kanade T (2014) Spatio-temporal event classification using time-series kernel based structured sparsity. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) Computer vision—ECCV 2014. Lecture notes in computer science, vol 8692. Springer, Switzerland, pp 135–150
16. Rezende DJ, Mohamed S, Wierstra D (2014) Stochastic backpropagation and approximate inference in deep generative models. ArXiv e-prints

17. Judd S (1988) On the complexity of loading shallow neural networks. J Complex 4 (3):177–192
18. Kingma DP, Welling M (2013) Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114
19. Kroener A, Haupert J, de Gea Fernandez J, Steffen R, Kleegrewe C, Schneider M (2013) Supporting interaction with digital product memories. In: Wahlster W (ed) SemProM. Cognitive technologies. Springer, Berlin, pp. 223–242
20. Lörincz A, Milacski ZÁ, Pintér B, Verö AL (2016) Columnar machine: fast estimation of structured sparse codes. Biol Inspired Cogn Architect 15:19–33
21. Mihaly A, Milacski ZA, Toser Z, Lőrincz A, Tettamanti T, Varga I (2015) Cost and risk sensitive decision making and control for highway overtaking maneuver. In: 2015 International conference on models and technologies for intelligent transportation systems (MT-ITS), pp 87–92. IEEE
22. Milacski ZÁ, Ludersdorfer M, Lőrincz A, van der Smagt P (2015) Robust detection of anomalies via sparse methods. In: Arik S, Huang T, Lai WK, Liu Q (eds) Neural information processing. Springer, Heidelberg, pp 419–426
23. Palotai Z, Lang M, Sarkany A, Toser Z, Sonntag D, Toyama T, Lörincz A (2014) Labelmovie: semi-supervised machine annotation tool with quality assurance and crowd-sourcing options for videos. In: 12th International workshop on content-based multimedia indexing, CBMI 2014, Klagenfurt, Austria, 18–20 June 2014. pp. 1–4. IEEE http://dx.doi.org/10.1109/CBMI.2014.6849850
24. Pistofidis P, Emmanouilidis C, Papadopoulos A, Botsaris P (2014) In: Proceedings of the European conference of the prognostics and health management society
25. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Networks 61:85–117
26. Schoeffmann K, Hudelist MA, Huber J (2015) Video interaction tools: a survey of recent work. ACM Comput Surv 48(1):14:1–14:34 http://doi.acm.org/10.1145/2808796
27. Sonntag D (2014) Ermed—towards medical multimodal cyber-physical environments. In: Schmorrow DD, Fidopiastis CM (eds) Foundations of augmented cognition advancing human performance and decision-making through adaptive systems. Springer, Heidelberg, pp 359–370
28. Sonntag, D (2015) Tutorial 3: intelligent user interfaces. In: Proceedings of the 2015 IEEE international symposium on mixed and augmented reality, ISMAR 2015. IEEE Computer Society, Washington, DC, USA, pp. xxxii–xxxii
29. Sonntag D, Toyama T (2013) On-body IE: a head-mounted multimodal augmented reality system for learning and recalling faces. In: 9th International conference on intelligent environments (IE), pp. 151–156
30. Sonntag D, Weihrauch C, Jacobs O, Porta D (2010) THESEUS usability guidelines for use case applications. Technical report, DFKI and Federal Ministry of Education and Research Germany
31. Sonntag D, Zillner S, Chakraborty S, Lőrincz A, Strommer E, Serafini L (2014) The medical cyber-physical systems activity at EIT: a look under the hood. In: 2014 IEEE 27th International symposium on computer-based medical systems, 27–29 May 2014. IEEE Computer Society, New York, pp 351–356
32. Tösér Z, Lörincz A (2015) The cyber-physical system approach towards artificial general intelligence: the problem of verification. In: Bieger J, Goertzel B, Potapov A (eds) Artificial general intelligence. Springer, Heidelberg, pp 373–383
33. Turunen M, Sonntag D, Engelbrecht K, Olsson T, Schnelle-Walka D, Lucero A (2015) Interaction and humans in internet of things. In: Abascal J, Barbosa SDJ, Fetter M, Gross T, Palanque PA, Winckler M (eds) Proceedings of the 15th IFIP TC 13 international conference on human-computer interaction Part IV—INTERACT 2015, Bamberg, Germany, 14–18 September 2015. Lecture notes in computer science, vol. 9299. Springer, Heidelberg, pp 633–636 (2015)
34. Šarić M (2011) Libhand: a library for hand articulation. http://www.libhand.org/, version 0.9

35. Wahlster W (2013) The semantic product memory: an interactive black box for smart objects. In: Wahlster W (ed) SemProM—foundations of semantic product memories for the internet of things. Cognitive technologies. Springer, Heidelberg, pp 3–21
36. Zhou A, Yu D, Zhang W (2015) A research on intelligent fault diagnosis of wind turbines based on ontology and FMECA. Adv Eng Inform 29(1):115–125

Applying Multi-objective Optimization Algorithms to a Weaving Machine as Cyber-Physical Production System

Marco Saggiomo, Yves-Simon Gloy and Thomas Gries

1 Introduction

Weaving is the most common as well as the oldest process for fabric manufacturing. Until today a fabric is created by crossing warp and weft threads in a right angle, like it was done since approximately 4000 B.C. Today's applications are for example:

- apparel (jeans, lining fabric, etc.),
- geotextiles (erosion protection, soil reinforcement, etc.) and
- technical textiles (filters, fireproof fabric, airbag fabric, reinforcements for fiber composites, etc.), see [9].

Because of the low production costs, the textile production has been relocated to the Asian countries, whereas the production of high-quality and technical textiles is progressively shifted to Europe. The textile industry in high-wage countries like Germany is facing numerous challenges today. For example, the tendency to small lot sizes requires shorter cycle times and aggravates the economical production of goods, see [4, 11].

M. Saggiomo (✉) · Y.-S. Gloy · T. Gries
Institut Für Textiltechnik der RWTH Aachen University (ITA),
Otto-Blumenthal-Str. 1, 52074 Aachen, Germany
e-mail: Marco.Saggiomo@ita.rwth-aachen.de

Y.-S. Gloy
e-mail: Yves.Gloy@ita.rwth-aachen.de

T. Gries
e-mail: Thomas.Gries@ita.rwth-aachen.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_20

Small lot sizes in the fabric production often involve a change of the fabric. A weaving machine with about 200 parameters has to be reconfigured after each change of the fabric to fulfill the expectations of the customer. In order to find the optimal configuration for the machine, the operator of the weaving machine has to conduct weaving trials. These time-consuming and wasteful trials require—depending on the experience of the operator—the weaving of up to 120 m of fabric until the optimal parameters are found, see [5].

This research paper presents an algorithm for multi-objective self-optimization of the weaving process and the integration of this algorithm into the machine control of the weaving machine. A weaving machine is upgraded to a cognitive unit (CPPS) on the shop floor. The algorithm for self-optimization identifies a combination of machine parameters, with the result that the essential objective functions can be adjusted to the individual preferences. Figure 1 visualizes the principle of a weaving machine.

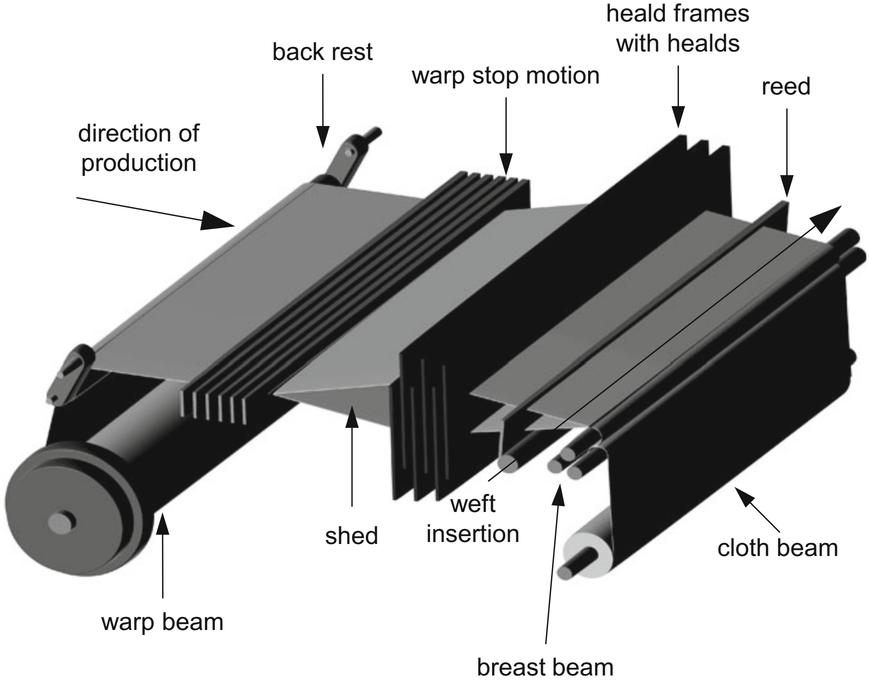


Fig. 1 Setup of a weaving machine [1]

2 Approach and Implementation of Multi-objective Self-optimization Procedure

Self-optimization systems apply adoptions of their inner state or structure in case of changes in input conditions or disturbances. Target values for self-optimization can be e.g. capacity, lot size, quality or energy consumption, see [4]. According to [7] self-optimization systems are characterized by the following continuous steps:

- analysis of actual situation
- determination of targets
- adaption of system behavior in order to reach the targets

The presented concept of self-optimizing production systems will now be applied to the weaving process. The following objective functions are considered by the multi-objective self-optimization (MOSO) of the weaving process:

- warp tension
- energy consumption of the weaving machine (air- and active power consumption)
- quality of the fabric

The objective functions are optimized according to the following parameters:

- basic warp tension (bwt)
- revolutions per minute (n)
- vertical warp stop motion position (wsm_y)

With the MOSO of the weaving process, a weaving machine is enabled to automatically find an optimal configuration. A program for self-optimization is implemented in a programmable logic controller (PLC). Figure 2 provides an overview of the required hard- and software infrastructure.

2.1 Signal Processing

For the signal processing and the execution of the self- optimization routine the ibaPADU-S Module system by iba AG, Fürth, Germany is used. The System consists of the following modules:

- ibaMS16xAI-20 mA: Analog input module for current signals in the range of [0…20] mA
- ibaMS16xAl-10 V: Analog input module for voltage signals in the range of [−10…10] V
- ibaPADU-S-IT-16: Central Processing Unit (CPU) for the modular system

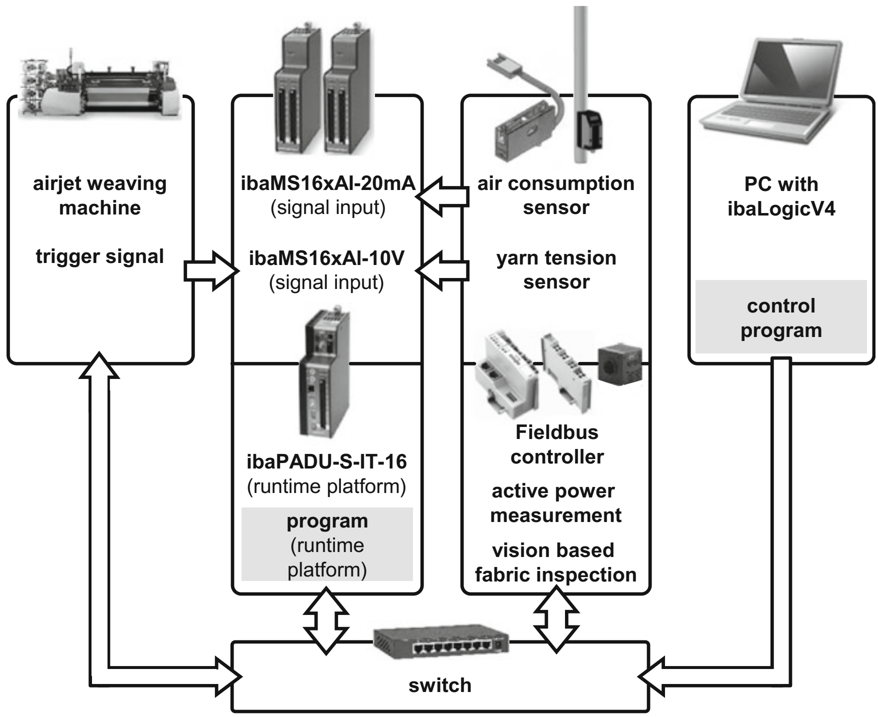


Fig. 2 Hard- and software infrastructure

The analog input modules ibaMS16xAI-20 mA and ibaMS16xAl-10 V collect and process the signals from the sensor system. Both analog input modules are connected to the base unit ibaPADU-S-IT-16 using a back panel bus. The base unit receives the data from the analog input modules through the back panel bus.

The central unit ibaPADU-S-IT-16 is connected to a computer using the Transmission Control Protocol/Internet Protocol (TCP/IP) interface. On the computer the software ibaLogicV4 from iba AG is installed. IbaLogicV4 is a programming environment which forms a software-based programmable logic controller (soft-PLC) together with the introduced modular system. An ibaLogicV4 program is created on the computer and transmitted to the central unit using TCP/IP. The central unit provides a runtime platform for the ibaLogicV4 program (runtime system). In the ibaLogicV4 program the data from the analog input modules are collected and processed. The program for MOSO is developed within the environment of ibaLogicV4 and uses the ibaPADU-S-IT-16 module as runtime system. The program for MOSO is presented in Sect. 2.4.

2.2 Measurement Technology of Warp Tension

For measuring the warp tension, the yarn tension sensor TS44/A250 by BTSR International S.p.A. Partita, Olgiate Olona, Italy is used. The yarn tension sensor generates a voltage signal in the range of [0…10] V which is proportional to the present yarn tension. The yarn tension sensor is placed in the middle of the weaving machine, between the back rest and the warp stop motion. The data connector of the yarn tension sensor is connected to the soft-PLC using an analog/digital converter. For additional information on the yarn tension sensor, see [8].

2.3 Measurement Technology for Energy Consumption

2.3.1 Air Consumption Measurement

The air consumption of the weaving machine is measured using the flow sensor SD8000 by ifm Electronic GmbH, Essen, Germany. The flow sensor generates a signal, which is proportional to the compressed air consumption in the range of [4…20] mA. The output data from the flow senor are wirelessly transferred to the soft-PLC.

2.3.2 Active Power Measurement

The power measurement module collects characteristic values of the three-phase supply and saves the values into the process image. To access the measurement values from the power measurement module, the power measurement module is connected to a Fieldbus controller using a terminal bus.

The process image of the power measurement module is provided to the fieldbus controller via the terminal bus. The fieldbus controller is connected to the soft-PLC using the Transmission Control Protocol/Internet Protocol (TCP/IP) Interface. The communication between fieldbus controller and soft-PLC is carried out in the Modbus-Protocol format. The soft-PLC sends out a specific request (Request) in the Modbus-Protocol formal to the fieldbus controller and receives the requested value from the process image (Response).

As soon as the response has been received by the soft-PLC, the requested data are available for the signal processing. Both, Modbus-Request and Modbus-Response, consist of binary codes and are organised as bytes. The runtime platform with soft-PLC is the Modbus-Master and sends the request to the fieldbus controller, which is the Modbus-Slave. The request contains information regarding the requested value from the process image of the power measurement module. In response to the request the fieldbus controller identifies the requested value and stores it into the response. The response is send via TCP/IP interface to the Modbus-Master. In the scope of this paper the active power is requested from the process image of the power measurement module.

2.3.3 Measurement Technology for Fabric Quality

At Institut für Textiltechnik der RWTH Aachen University (ITA), Aachen, Germany a measuring system for online error detection during fabric production was developed [12]. A camera takes pictures of the fabric. Subsequently the pictures are checked for defects in the fabric, using digital image processing. The software for digital image processing runs on a separate computer. The camera system is installed over the section of the weaving machine where the fabric is created.

The camera system is able to detect defects immediately after the fabric is produced. The digital image processing software is calibrated using a flawless piece of fabric. The digital image processing classifies deviations from the calibrated condition as a defect [12]. Depending on the share of incorrect pixels in the pictures, the fabric is assigned to a quality category.

The examination for defects is carried out in real time during the weaving process. The computer running the digital image processing is connected to the soft-PLC via TCP/IP interface. Depending on the status of the fabric, the number of the quality category (0–4) is continuously transmitted via TCP/IP. A quality category of 0 is achieved in case the fabric quality is accurate. A quality category of 4 stands for a destroyed fabric.

2.4 *Program Steps*

The program for MOSO consists of the steps shown in Fig. 3. Though continuous communication between weaving machine and soft-PLC, the weaving machine is enabled to run the entire program autonomously.

In the first step an experimental design is calculated automatically. Within this design, the three setting parameters static warp tension, vertical position of warp

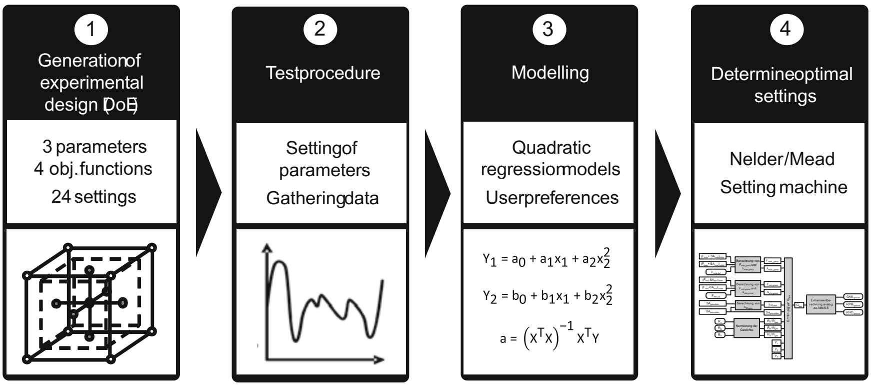


Fig. 3 Program steps of multi-objective self-optimization of the weaving process

stop motion and revolutions per minute are varied. The user sets the parameter spaces to ensure that the algorithm acts within a feasible range.

During the second step, the test procedure, the weaving machine sets-up every test point. Sensor data describing the objective functions are recorded for the respective parameter setting.

In the third step, the obtained data are used to calculate three regression models (one model per each objective function) which describe the objective functions in dependence of the setting parameters.

In the last step, an optimized set-up of the weaving machine based on predefined quality criteria is calculated by application of desirability functions and a numerical optimization algorithm. Before execution of the optimization procedure, user-defined preferences regarding the objective functions (warp tension, energy consumption and fabric quality) can be integrated through target weights. The preference scale for each objective function is divided into three sections (low, middle, high).

The program for MOSO is implemented within the ibaLogicV4 programming environment and runs on a central processing unit as depicted in Sect. 2.1.

The next chapter illuminates desirability functions and the optimization algorithm used for MOSO.

3 Desirability Functions and Nelder/Mead Algorithm

3.1 Desirability Functions

The origin of the application of desirability functions in the multi-dimensional optimization goes back to Derringer and Suich [6]. The aim of using desirability functions is to summarize the objective functions which need to be optimized into one common function. The aggregation of the objective functions is conducted using the so-called desirability. For each objective function, one desirability function is developed. The desirability function assigns a desirability to each value of the objective functions. The desirability function has a value range of [0; 1]. If the value of one objective function reaches a desirability of zero, the result is invalid within the optimization routine. In case the desirability reaches the value one, the value of the objective function is optimal. The desirability w_Z is plotted over the normalized objective function Z(X). Desirability functions can be constructed in three different ways. If the goal is to achieve the highest possible value for one objective function, the desirability function for maximizing has to be used. The desirability increases when the objective function value increases, etc.

The aim of the utilization of desirability functions is to aggregate the target functions into one common function, the so-called total desirability d_{tot}. d_{tot} is calculated by using the geometric mean of the individual desirabilities:

$$d_{tot} = (w_1, w_2, \ldots, w_n)^{\frac{1}{n}} \quad (1)$$

whereas w_1, w_2, …, w_n are the desirabilities of n objective functions.

The total desirability reveals how close the individual desirabilities are to the optimal range. Because of the multiplication of the individual desirabilities, dtot is in the range of [0; 1]. A total desirability of one is reached, when all target functions are in the optimal range. In case only one target function has an invalid value, the total desirability equals to zero.

The combination of process parameters which maximizes d_{tot}, represents the optimal operating point for the weaving process.

Numeric algorithms are suitable for maximizing the total desirability. The application of numeric algorithms is more efficient than for example grid search methods [3]. It is advised in several references to utilize Nelder/Mead algorithm [10] to maximize the total desirability, see e.g. [2].

3.2 *Nelder/Mead Algorithm*

The Nelder/Mead algorithm is a numeric optimization procedure [10]. To find a subjective optimal operating point of the weaving machine, d_{tot} is maximized. The Nelder/Mead algorithm searches for a combination of the three parameters basic warp tension, revolutions per minute and vertical warp stop motion position that maximizes d_{tot}.

Setting the start values for the considered parameters leads to the starting point for the algorithm. The start values are set before the first iteration and are moved towards the optimal values during the utilization of the algorithm. Starting from a minimization problem with m parameters, the algorithm considers m + 1 parameter combinations (P_1, P_2, …, P_{m+1}). The values of the objective functions are calculated in the m + 1 points and sorted ascendingly. The next step is the examination of the three points P_1, P_2 and P_3 in the parameter space. At each of the three points the algorithm calculates the value of the objective functions $F(x_1,x_2)$:

$$F_i = F(P_i),\ i = 1, 2, 3 \quad (2)$$

Afterwards the function values are sorted:

$$F_1 \leq F_2 \leq F_3 \quad (3)$$

Considering this example, F_3 is the worst (highest) and F_1 is the best (lowest) value in the context of the optimization. The minimization of the target function is achieved by applying several iterations of the algorithm. In each iteration one new point in the parameter space is created, which replaces the point P_{m+1} with the biggest value of the function to be minimized. In the present case the point P3 results is the worst value of the target function and is therefore replaced in the next iteration. The replacement of the worst point is achieved through the basic operations of the Nelder/Mead algorithm.

3.3 Experimental Results

In this chapter, the ibaLogicV4 program for MOSO of the weaving process is validated during a long-term test in the laboratory of ITA. To establish industrial conditions, the duration of the long-term test is eight hours, like usual shift duration. A long-term test is carried out using the MOSO against not using the optimization procedure respectively, to examine the influence of MOSO on production figures. The long-term test is conducted with an air-jet weaving machine OmniPlus 800 by Picanol n.v., Ieper, Belgium. During the long-term test, polyester filament yarn with 330 dtex for warp and weft was used (binding: twill 3/1). The configuration of MOSO used for the long-term test are listed in Table 1.

After program execution of MOSO, the algorithm calculates the following optimal parameter settings: bwt = 3,71 kN; n = 522 RPM; wsmy = 20 mm.

The following settings are used as reference settings coming from an industrial weaving mill that processes the same material as mentioned above: bwt = 4 kN; n = 900 RPM; wsmy = 0 mm. During the long-term test the following parameters of the weaving process are recorded:

- efficiency of the weaving machine
- produced amount of fabric
- amount of weft insertions
- warp/weft defects and breakages

Additionally, data of the objective functions are recorded. The results of the long-term test using MOSO and reference settings are shown in Table 2.

During the long-term test, sensor data regarding the objective functions are recorded using the software ibaPDA from iba AG, Fürth, Germany. The measured

Table 1 Configuration of MOSO used for long-term test

Setting	Value
Lower/upper limit bwt	2 kN/4 kN
Lower/upper limit n	400 RPM/900 RPM
Lower/upper limit wsm_y	0 mm/20 mm
Target weight warp tension	Low
Target weight energy consumption	Low
Target weight quality	High
Algorithm start point bwt/n/wsm_y	3,5 kN/750 RPM/15 mm

Table 2 Results of long-term test

Recorded data	Results	
	MOSO	Ref. settings
Efficiency (prod. time/total time)	98.6 %	97.2 %
Produced fabric	8.16 m	15.41 m
Weft insertions	125,157	215,982
Weft defects	2	6
Warp breakages	0	0
Average warp tension	1.27 N	1.49 N
Average air consumption	134.23 m^3/h i. N	155.26 m^3/h i. N
Average active power usage	2.49 kW	4.62 kW
Average quality category	0.93	1.55
Set-up time	30 min	120 min

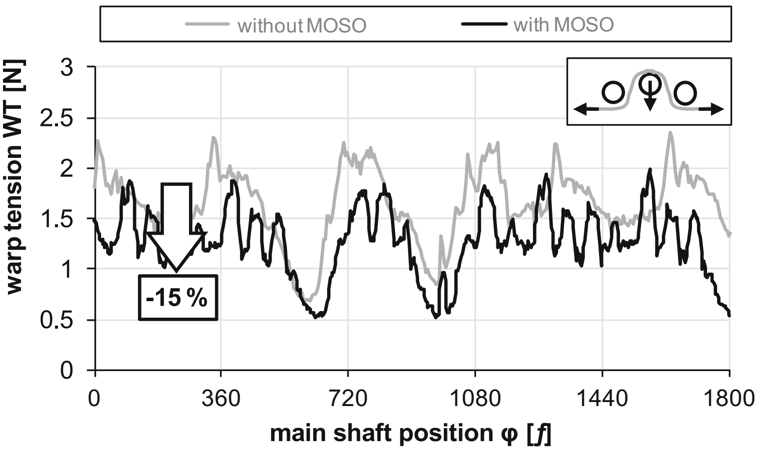


Fig. 4 Comparison of warp tension with and without MOSO

data is illustrated in Figs. 4, 5 and 6. Data are plotted over the main shaft position of the weaving machine which is the rotating angle of the machine's main drive.

The program for self-optimization enables the weaving machine to autonomously find an operating point, which improves all objective functions compared to conventional (reference) machine settings.

The efficiency presents the relation of production time of the weaving machine to the total time. The efficiency of the weaving machine is higher using the optimal setting than in case of using the reference settings, see Table 2. Higher efficiency is mainly achieved by reduced machine downtime.

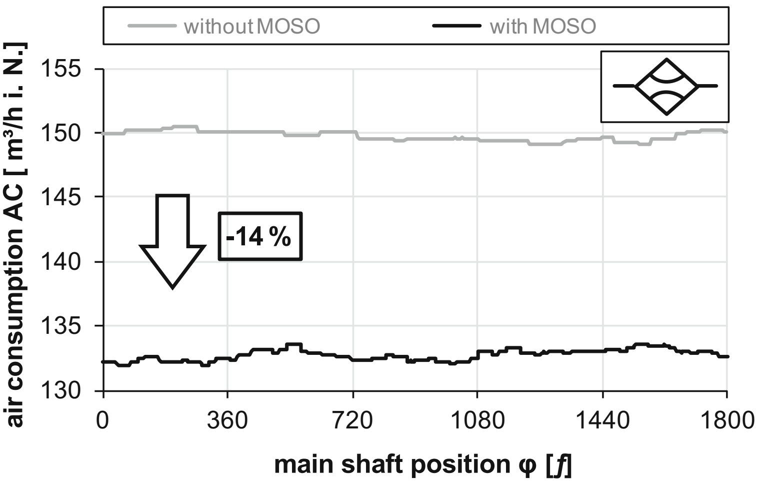


Fig. 5 Comparison of air consumption with and without MOSO

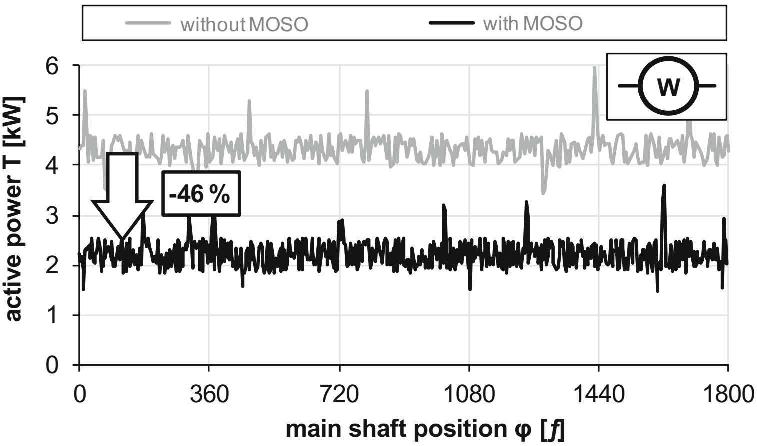


Fig. 6 Comparison of active power consumption with and without MOSO

Using the optimal machine settings, two weft defects caused be the collision of the weft threads with sagging warp threads occurred. In contrast, using the sub-optimal settings, six weft defects occurred. During the long-term test it was observed, that the machine runs more stable with less RPM. The higher amount of weft defects can be explained by a disadvantageous machine speed of 900 RPM. Weft defects result from the faulty transport of weft threads across the width of the weaving machine.

Without MOSO a machine operator needs around 120 min for the configuration of the weaving machine and to find appropriate settings for the process. The program for self-optimization is concluded in 30 min and successfully reduces the set-up time by 75 %.

4 Conclusion and Outlook

This paper presented a concept and implementation of multi-objective self-optimization of the weaving process. An optimization routine which is implemented into a software-based programmable logic controller enables the weaving machine to calculate optimal parameter settings autonomously. Individual preferences of operators or plant management are integrated into the calculations. By assistance of the resulting cognitive weaving machine, the set-up time was reduced by 75 %.

Further research will focus the development of mobile applications to form an assistance system for operators basing on the presented results. Moreover, the cognitive weaving machine will be embedded into an intelligent textile process chain in the sense of Industry 4.0 where all production units are interconnected to each other.

References

1. Adanur S (2001) Handbook of weaving. CRC Taylor and Francis, Boca Raton
2. Bera S, Mukherjee I (2010) Performance analysis of nelder-mead and a hybrid simulated annealing for multiple response quality characteristic optimization. In: Proceedings of the international multiconference, vol 3, pp 1728–1732
3. Blobel V, Lohrmann E (1998) Statistical and numerical methods for data analysis, vol 1. Vieweg+Teubner, Wiesbaden. Original title: 'Statistische und numerische Methoden der Datenanalyse'
4. Brecher C (ed) (2011) Integrative production technology for high-wage countries. Springer, Berlin, pp 747–1057
5. Chen M (1996) Computer-based optimization of the weaving process according to warp thread stress and warp processing behaviour. Dissertation, Universität Stuttgart, Stuttgart. Original title: 'Computergestützte Optimierung des Webprozesses bezüglich Kettfaden-beanspruchung und Kettlaufverhalten'
6. Derringer G, Suich R (1980) Simultaneous optimization of several response variables. J Qual Technol 12(4):214–219
7. Gausemeier J, Rammig FJ, Schäfer W (2009) Design methodology for intelligent technical systems: develop intelligent technical systems of the future. Springer Science and Business Media, Heidelberg
8. Gloy Y-S, Sandjaja F, Gries T (2015) Model-based self-optimization of the weaving process. J Manuf Sci Technol (CIRP) 9:88–96
9. Gries T, Veit, D, Wulfhorst B (2015) Textile technology. Carl Hanser, Munich

10. Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7(4): 308–313
11. Osthus T (1996) Process optimization and changeover time reduction for weaving through automatical adjustment of backrest and warp stop motion. Aachen, Rheinisch Westfälische Technische Hochschule Aachen, Dissertation. Original title: 'Prozessoptimierung und Rüstzeitverkürzung in der Weberei durch automatische Einstellung von Streichbaum und Kettwächterkorb'
12. Schneider D, Gloy Y-S, Merhof D (2015) Vision-based on-loom measurement of yarn densities in woven fabrics. Instrum Measur IEEE Trans Instrum Measur 64(4):1063–1074. doi:10.1109/TIM.2014.2363580

Cyber Physical Production Control

Autoren G. Schuh, V. Stich, C. Reuter, M. Blum, F. Brambring, T. Hempel, J. Reschke and D. Schiemann

1 Current Challenges of Production Control

Today, manufacturing companies are increasingly confronted with the influences of a dynamic environment and the ensuing of continuously increasing planning complexity. Companies have to adapt themselves and their processes to dynamic environment conditions like movements in customer demand, reschedules in supply as well as turbulences in networks. Nevertheless, a successful production control is characterized by high process efficiency and a high availability of information.

V. Stich · M. Blum · J. Reschke · D. Schiemann
Institute for Industrial Management (FIR), RWTH Aachen University, Campus-Boulevard 55, 52074 Aachen, Germany
e-mail: Volker.Stich@fir.rwth-aachen.de

M. Blum
e-mail: Matthias.Blum@fir.rwth-aachen.de

J. Reschke
e-mail: Jan.Reschke@fir.rwth-aachen.de

D. Schiemann
e-mail: Dennis.Schiemann@fir.rwth-aachen.de

A.G. Schuh · C. Reuter · F. Brambring · T. Hempel (✉)
Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Steinbachstr. 19, 52074 Aachen, Germany
e-mail: T.Hempel@wzl.rwth-aachen.de

A.G. Schuh
e-mail: G.Schuh@wzl.rwth-aachen.de

C. Reuter
e-mail: C.Reuter@wzl.rwth-aachen.de

F. Brambring
e-mail: F.Brambring@wzl.rwth-aachen.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_21

Examples for these challenges are steadily decreasing process times, increasing product and process variations as well as the unclear market and production environment [36, 42, 52].

In this context, delivery times in the machine and plant manufacturing industry have been reduced by about 50 %, which has massive influence on the capacity flexibility and also on the entire order fulfillment process [48]. An example of the realignment of the order fulfillment process is found at the company *Siemens*. In the healthcare business unit the processing time of a computer tomography scanner has been reduced by 67 % and the delivery time by about 86 % through a consistent logistical orientation [38]. According to Schuh, the product and process complexity has reached such a high level that it can only be controlled under very high monitoring and coordination efforts [28]. Other challenges for the production planner are the volatile economic situation, fluctuating customer demands, short-term rescheduling, an early release of orders or capacity overload [44]. Furthermore, incorrect or outdated data often leads to poor planning results. The production planner often has no appropriate evaluation variables for control decisions. With several production planners on duty, several locally optimal decisions can exist, which can have a negative effect on global corporate targets [23].

Today the granularity of different production data is often not sufficient. As numerous practical examples from projects with the industry have shown in the last years, the typical overall throughput time of a production process is composed as follows: The production process time is about 5 % whereas the transition periods are about 95 % of the overall throughput time. Transition periods are the time intervals between the last process step of a previous machine and the first process step of a following machine. Concerning the production processing time plenty of high resolution information is available in many businesses, however, transition periods are often mapped only rudimentary without much information [27].

Concerning the above-mentioned, upcoming challenges in the corporate environment, the production control is especially facing issues of low forecast quality in production planning. Due to uncertainties of future market requirements, unpredictable machines failures and unreliable suppliers, the forecast quality is inaccurate. The current approaches to solve this dilemma are a variety of IT-systems which, for the most parts, do not allow reliable forecasts for the planning and control due to inaccurate dispositional data [26, 30, 33].

In order to counteract these shortcomings, the approach of a cyber-physical production control follows the idea of implementing methods from the field of data analytics to production control processes. In this chapter, current research directions as well as implications for a future production control are outlined.

Driven by the digitalization in the context of *Industrie 4.0*, the importance of data processing is steadily increasing. Although the automation of machines and the efficient organization of the shop floor are still success factors to increase productivity, operational data and its intelligent usage is becoming increasingly important. To enable the intelligent usage of data, technologies from the field of data generation, data processing, data evaluation and data exchange need to be implemented. The generated data alone has no real benefit for companies. First the intelligent transformation provides additional value by allowing conclusions about past events and predictions about future conditions of the production. By means of self-learning methods, the industrial business applications support the user. With the help of the so called smart data, a comprehensive transparency of the whole process chain is possible, planning accuracy can be increased and the quality of managerial decisions can be improved [1, 46, 50].

2 Vision of a Cyber Physical Production Control

From the descriptions above, one can assume that there are numerous difficulties for today's production planners and controllers, which cannot be solved adequately with existing tools. Possible solutions to the dilemma described in the first section are new adaptive prediction and control systems, supporting the production controller and planner in preparing for the rising challenges.

This section shows which solution approaches exist for similar challenges in other professions in order to conclude possible new solutions for production control.

2.1 Smart Decision Support in Daily Life

Intelligent decision support systems of our everyday life are, for example, intelligent weather applications for smartphones or tablets. The app *wetter.com—Wetterwecker* knows the current and future weather at the current location of the app user with the help of the GPS module. The alarm clock adjusts its preset time according to the weather development overnight. With heavy cloud cover it wakes you five minutes earlier, when it rains 15 min and in times of new snow 30 min earlier. The tolerance can be manually adjusted. With this app, there is no need to be afraid to miss an important morning meeting by a surprising onset of winter [49].

Another example for intelligent decision support comes from the car manufacturer Rover, with its use of the head up display technology. Besides key driving information, which can be displayed on a head up display, like road speed, gear position, turn-by-turn navigation, traffic sign recognition, the cruise control set speed including warnings, Rover offers a windshield-wide virtual windshield. With the help of a central lasersensor it projects the parking trajectory on the virtual windshield and enables an assisted, safe parking process. Another feature, especially useful in the racing sport, is the visualization of an ideal racing line or the last driven lap on a racetrack on the virtual windshield. Best possible braking points or peaks can also be highlighted and assist a driver in reaching minimum times on a racetrack [11]. With the help of this technology, a race driver can reach those minimum times much faster than without that innovative decision supportive virtual windshield. An example application is shown in Fig. 1.

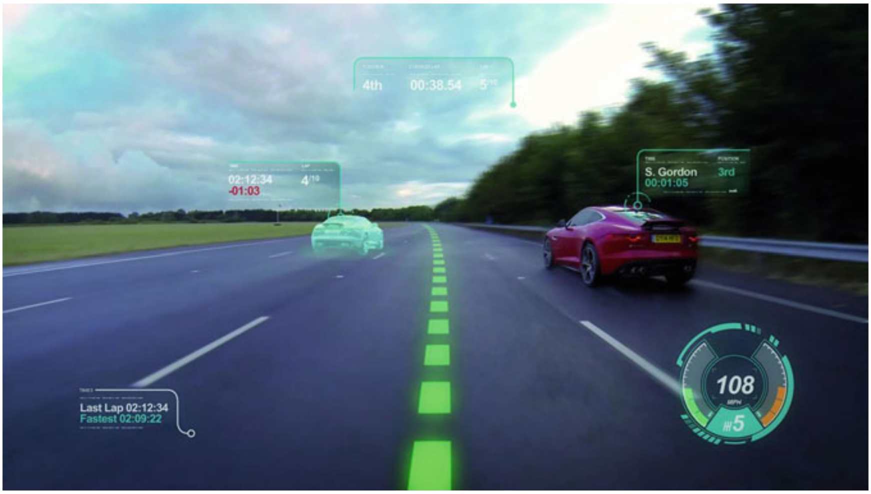


Fig. 1 Virtual windshield by Rover/Jaguar [11]

These two illustrated examples of intelligent decision support systems in non-production life represent how a prediction with a certain probability can be made with intelligent linking of collected data for situations, which are difficult and hard to overlook for human beings.

Applied to production it would be of great benefit if the production planner could increase his decision-making and confidence level by implementing a simple real-time visualization and interpretation of the current and expected production events. The enablers for that vision in production are stated in the following section.

2.2 Enabler for Decision Support Systems in Production Control

In the following, crucial enablers for a successful operation of a cyber physical production control are stated. The crucial enablers are Cyber-Physical Systems (CPS), fast mobile internet infrastructure and Automatic Model Generation in Simulation Software.

2.2.1 Cyber-Physical Systems

The term "cyber-physical" includes the most important characteristic of this concept: the combination of embedded systems to monitor physical processes within a digital network [4, 21]. A variety of sensors and actors build a networked basic structure of control and regulation tools for close to real-time data processing, which can take decisions autonomously and adapt to new surrounding conditions to a certain extent. This type of system is already used in many applications, for example in mobile phones, modern aircrafts or nuclear power plants and many other areas [4, 21].

On the one hand, the challenges exist in the targeted configuration and further development of existing systems with their large number of required characteristics and on the other hand in the development of additional application areas. Due to their almost unlimited application possibilities, CPS are expected to be controllable and ready to use in all application areas and to facilitate taking full advantage of their economic and social benefits [4, 21].

The massive use of CPS in future production environments clears the way for building the so called Smart Factory. Such a Smart Factory will not be deserted. In a smart factory individual customer orders are controlling and triggering production processes as well as the supply chain [2]. But the employees are carriers of decisions and carry out important functions in design, installation, conversion and maintenance of increasingly complex and networked cyber physical production systems. These new requirements in the context of Industrie 4.0 require new assistance systems and multimodal user interfaces with the production process, the machinery and equipment as well as software systems involved.

The digital processing of CPS can be obtained by sensors, actors, identification and localization technologies, information processing and communication technologies. The digitally enhanced production participants can then be considered as Smart Objects and build the basis for Smart Products, Smart Data or Smart Services [20].

Smart Objects are objects, which in terms of embedded technologies are aware of their environment and the conditions. They may be able to make decisions independently, to convey actual information and to perform or stimulate actions. So they are combining physical objects with data processing. Smart Objects, as described above, offer completely new possibilities for upcoming productions through their abilities to identify, communicate, process information, save data. A network-based, full transparency of the current development and production status is made possible by massive digital finishing of the production. This in turn can increase the statement- and decision-making capability of the production manager enormously [22, 24].

2.2.2 Fast Mobile Internet Infrastructure

Another major enabler for cyber physical production control will be a global fast mobile internet infrastructure, which will replace rigid local BUS and WiFi networks in production.

In Fig. 2, the development of the transfer rates of wireless LAN technologies as well as mobile internet is shown. Wireless LAN started at around 1 MBit/s in 1996, whereas GPRS started in the late 1990s with a maximum of 55 kb/s. Current LTE transfer rates allow around 100 MBit/s, while the latest wireless LAN technology

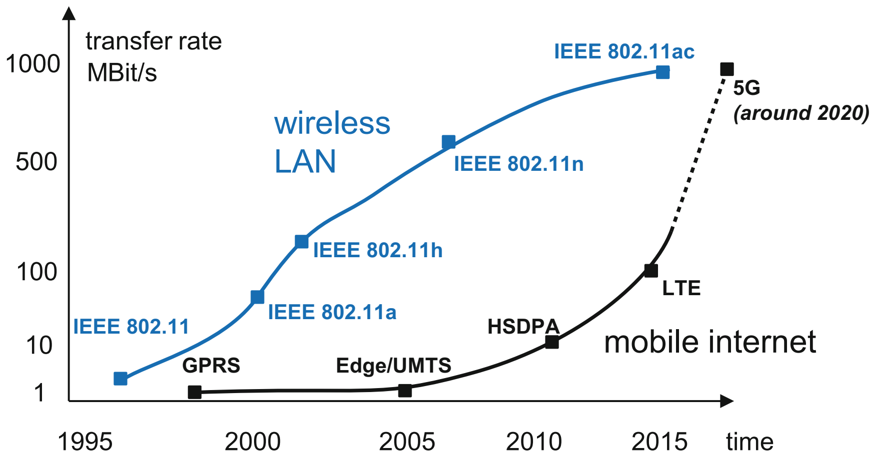


Fig. 2 Transfer rates of wireless LAN and mobile internet technologies (own illustration)

can achieve a maximum of about 1000 MBit/s. Over the last years, the curves of the transfer rates of WLAN and mobile internet are converging. If the trend is following the prediction, mobile internet of the fifth generation will outplace wireless LAN within the next few years.

By making use of those—now fast enough—mobile internet technologies for communication between machines, machines and software as well as machines and humans, companies become a lot more networked and agile [18].

2.2.3 Automatic Model Generation in Simulation Software

The application of automatically generated models of the manufacturing systems is an approach for facing the demand for continuous model adjustments used for planning in order to represent the real situation on the shop floor. This so called automatic model generation was developed to accelerate the generation process and enhance the accuracy of the simulation models, which is crucial for deriving decision support based on the generated models [16, 38].

Two exemplary approaches for the automatic model generation are given by Selke and Kapp [19, 39]. In their approaches, as well as in general, automatic model generation is based on planning and feedback data from production. Data for the model can be generated manually via Production Data Acquisition (PDA) solutions or derived automatically from the machines via Machine Data Acquisition (MDA) surfaces. MDA is preferable due to several reasons: it realizes an automatic interface between the machines and the accompanying IT-system and it can also easily track disturbances (like machine breakdowns). Major objective of the automatic model generation is to derive valid simulation models which are close to the real situation in production. Most automatic model generation approaches verify and validate their simulation model with a comparison to real feedback data.

After the description of the major technological enablers for a cyber physical production control in this section, the next section will focus on the use and processing of data to enable cyber physical production control.

3 Data Analytics Enable Cyber Physical Production Control

General questions of the production controller like *“What is the current status of my order?”* or, *“When can I deliver the order?”* are often very difficult to answer because of inconsistent and incorrect data with the dominating IT-system environment in production. This is becoming an increasing problem when automatic correction mechanisms are missing [31].

There are numerous support systems for production. Some essential systems include:

- Enterprise resource planning systems (ERP; holistic, process-oriented software solutions, which control and evaluate the commercial and technical operation rules) [37]
- Production planning and control systems (PPC; scheduling, capacity and quantity-based planning and control of manufacturing and assembly processes) [10]
- Manufacturing Execution Systems (MES; contemporary support of all production-relevant business processes of an enterprise) [45].

Advanced Planning and Scheduling (APS) systems, which are an evolution of PPC systems, represent an approach, by means of mathematical models and improved data, to achieve a more accurate prediction of future production statuses [14, 41].

As described above, APS systems reduce the gap between the planning and the current status, but they are not able to completely fill the gap. The calculated results are based on certain assumptions and are highly dependent on the data quality. For example, if confirmed data is inconsistent or even incorrect, an APS system fails [34]. Moreover, they have the disadvantages of strong turbulence and a lack of transparency of the planning algorithms so that acceptance of these systems by production planners is rather low [13].

The deficiencies in existing forecasting systems make it necessary to take action, since e.g. the matching of planning and real situation (planning quality) of many systems is as low as 25 % after only three days [16]. For a production system which shows complex system dynamics and is subject to highly diverse influences, a reliable production forecast would be of great help. With such a forecasting system, it would be easier to answer the essential questions above.

The application of high quality data for analytics has historically been much higher in the retail business than in the manufacturing industry. Therefore, traditional industries can benefit from the already developed and applied methods which are an ideal input for innovative solutions. The company Blue Yonder, which is Europe's leading Software as a Service (SaaS) provider for predictive applications, mainly for the retail industry, states the following vision: "Data shapes our lives, how we communicate, work and do business. With Predictive Analytics, we are revolutionizing the way that we all deal with data. We want organizations to gain insights from their data and be able to turn this information into business processes. We help them to become predictive enterprises that not only master digital change, but also shape it—innovatively and with a focus on the future." [3] The derived approach of the data analytics in production environments is described in the following (see Fig. 3).

A major prerequisite for a Cyber Physical Production Control is the digitization of production which aims to create a digital shadow of production processes in the accompanying IT-systems. Digital shadow is used in this section for the abstract image of the production process and its organizational flows in production related IT-systems. The digital shadow is necessary in order to generate data which can be used for data analytics. The goal of this first phase is defining necessary data sources, making the data measurable and understandable for information technology and interlinking all those sources of information.

After the first phase of the digitization of production and its processes, an existing business intelligence approach for data analytics [9] has been adapted to production, which consists of 4 successive steps (see Fig. 4):

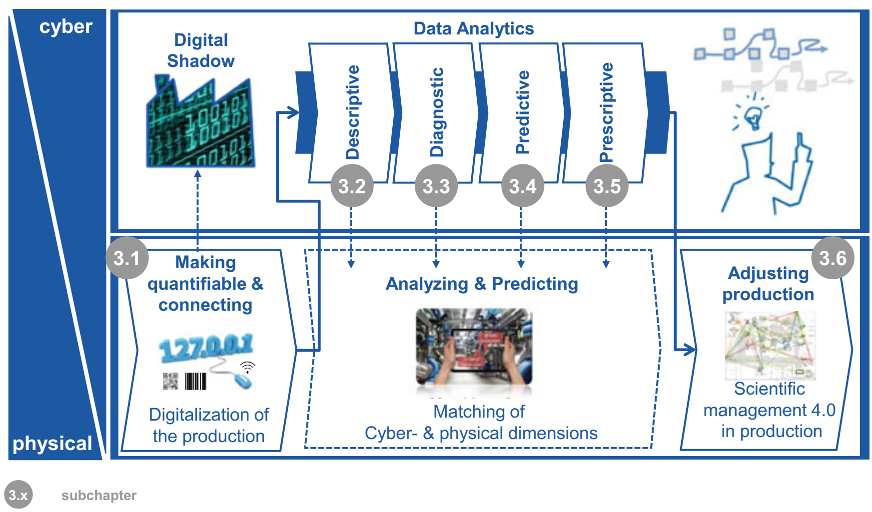


Fig. 3 Data analytics as the enabler for cyber physical production control (in dependence on Evans [9])

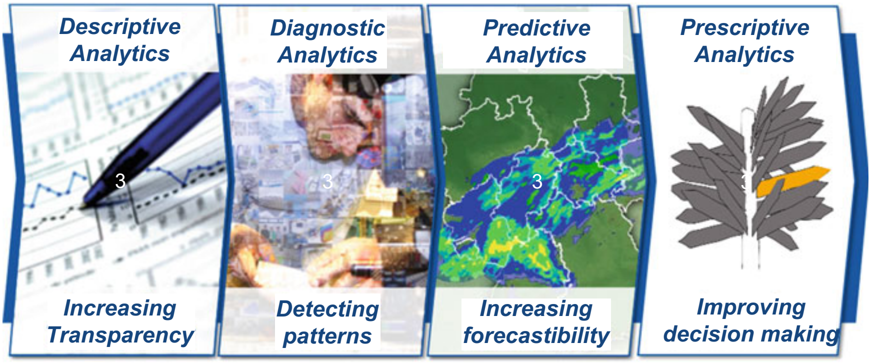


Fig. 4 Data analytics process (in dependence on Evans [9])

- In the first step, the *descriptive analytics*, the focus lies on increasing transparency of the production through the generation of data. In order to control and optimize production, it is extremely important to obtain a digital shadow. This is only possible through massively equipping of production with sensor technology of any kind, such as barcode, radio-frequency identification (RFID), camera, indoor GPS etc.
- In the second step, the *diagnostic analytics*, recognizing patterns within the generated data is the major objective. Recurring patterns in collected data can e.g. supply information about repetitive machines sequences in the production process or seasonal variations within a production cycle.
- In the third step, the *predictive analytics,* the goal is to build forecasts with the help of the previously identified patterns. Forecasts in everyday life are already integrated in the form of weather forecasts or the rear view camera of a car. In production, forecasts in the form of a prediction, when, for example, a bottleneck situation is expected or when a machine would probably break down are not really spread far yet [35]. Predictive Analytics offers great potential for production planning in order to obtain a smooth production process.
- The last step of the data analytics approach is the *prescriptive analytics*. It is the base for generating decision support for many managerial problems and questions and therefore enables quantified and reasonable decisions in todays and tomorrows demanding business surrounding.

At this point, it should be noted again, that a correct and up-to-date status of the production process in the IT-systems is of utmost importance, in order to generate reasonable decision support for optimizing the real production. Therefore, a constant check for congruence of the "physical" shop-floor level and the "cyber" digital shadow level has to be performed. The last step of the approach is the actual execution of the proposed actions by the personnel in charge, in order to optimize the production. In the following sections, the approach is described in broader detail.

3.1 Data Quality as an Enabler of Cyber Physical Production Control

A high-resolution data base acts as basis for the different steps of data analytics [43]. Thus, the data has to be correct, up-to-date, consistent, complete and redundancy-free [17]. Although companies mention the importance of data recording, a high-resolution and consistent data quality is still not self-evident. This is shown by the results of various surveys which state that only about 19 % of all surveyed companies believe that their data is being maintained in good quality and consistency [5]. The current bad data quality leads to an increasing importance of the data management for companies, especially among value added aspects [12, 47]. To successfully implement concepts of data analytics all relevant data of a product have to be gathered. The data base consists of many different data types like process times, location data, inventory date and many more (see Fig. 5).

To generate the digital shadow, a sufficient data quality is needed. According to Hildebrandt data quality is therefore the central factor of success for all logistical applications [17]. The three “V” of Big Data (*Volume, Velocity, Variety*) strongly influence the approaches of *Cyber Physical Production Control* [36]. To fulfill the requirements of forecast and learning ability, it is necessary to focus on the following five sectors of data quality.

By means of intelligent sensors which result in a higher amount and quality of feedback data from production, an increase in transparency of the current state of production can be obtained. The increased number of data corresponds to the first “V” (*Volume*) of Big Data and requires a useful handling of the grown amount of data [36]. The food manufacturer *Zentis* provides an example of improved data quality. *Zentis* implemented RFID technology in its container management and therefore improved its data situation: Better planning data, comparison of stock

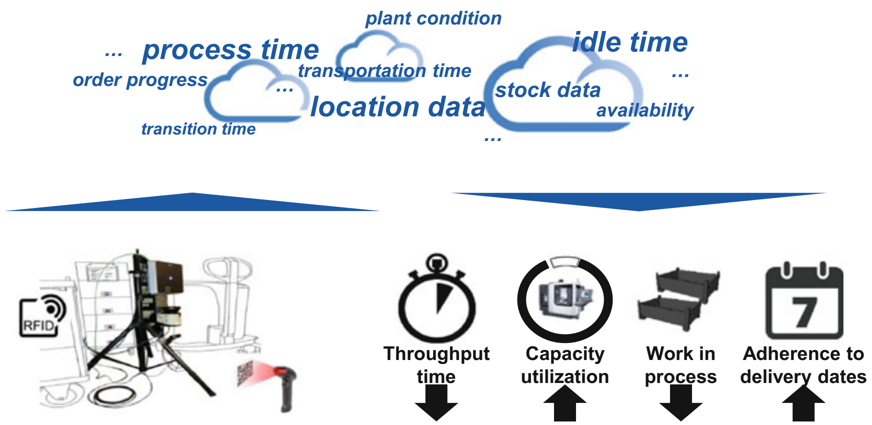


Fig. 5 A high-resolution data base as a base for data analytics

levels and customer needs, as well as quality improvements are realized benefits [7]. By doing this, a real-time monitoring of status and content is possible. On the one hand, this enables condition monitoring as well as the determination of location and position (e.g. important for liability issues in case of container whereabouts). On the other hand, the management of empty containers and return logistics is being enhanced.

The St. Gallen company *Intellion* chooses a different approach using RFID technology: RFID tags were attached to the front of storage boxes in the warehouse of a subsidiary at the fastening technology group *SFS* in Heerbrugg. As soon as the stock in the box is running low, a warehouse employee has to rotate the box by 180 degrees and the system places an order automatically [40]. As in the case of *Zentis*, a manual entry in the production support system is no longer necessary. By minimizing error-prone manual entries of the staff, the data quality increased radically.

In order to improve the data quality, inconsistencies in the returned production data can be automatically completed and repaired by the use of probability functions. To show the influence of inconsistent data the system can run two simulations: In the first run, it is assumed that the available data is fully consistent. Afterwards, the results of this simulation are being compared to the results of a second simulation, in which the failure of a machine is being simulated. This way, the influence of data inconsistencies can be determined and set as a control variable in the control circuit [34].

Regarding the high manual effort to maintain e.g. stocks, new technologies need to be implemented to simplify this process and improve the consistency of real-time information in the planning and control systems. By means of integrated camera on container level, such as the Würth iBin-container, it is possible to detect and automatically transmit information like the filling level, numerical and order information to the ERP system via RFID technology [51]. Therefore a consumption based delivery of small parts for the production via real-time transmission including an image is possible. Scales may also provide information about the content of a container, since they can track material withdrawals through a difference in the weight of a container. These again communicate through the network directly with the ERP system and can place an order at the right time. This mainly concerns to the second "V" of Big Data, *Velocity*, which refers to the increased respond speed by the system.

To obtain more accurate results for transition times in the planning systems and to improve data quality, heuristics and past values can be used which replace the average time data stored in systems by increased accuracy. By now most companies use values for the transition times which rely on average values. Incorporating the idea of a control circuit, average values cannot be used as a control variable. This is the reason why reliable transition times, which are based on past values, are necessary. This mainly concerns the third "V" of Big Data, *Variety*, which describes the width of the data base, caused by the available information objects.

The German automotive supplier *Westaflex* uses high-resolution production feedback data for example to monitor tooling machines and to enhance the controllability of their production system. This allows forecasting the tool wear and furthermore the product quality. As a result, preventive maintenance activities can be initiated [29].

3.2 Descriptive Analytics

One of the first areas regarding the use of data analytics is the area of descriptive analytics. Descriptive analytics aims at analyzing large collections of data from the shop-floor as well as data stored in the ERP- or ME-System with the purpose of getting an insight and conclusion of what happened in the past. To answer the main question of what happened in the past the data is analyzed by means of descriptive statistics. In order to efficiently aid decision making the data is analyzed by pre-defined criteria to find deviations from the norm. In the environment of the production system interesting parameters include processing times, transition times or stocks in the warehouse (see Fig. 6). By means of descriptive statistics, the system can automatically detect deviations and thus enables significant improvements for the company.

The benefits of the first step of data analytics are characterized by an increased transparency and reduced labor and inventory costs. By the use of key performance indicators descriptive analytics allow details about the frequency of events, labor costs, delivery times etc. Using dashboards or reports employees are provided with the right information. These views may differ depending on the target group addressed by the level of detail and its degree of aggregation. The production planner needs to keep an overview of all processes, but e.g. a machine worker does only need specific information like working plans or shift information.

Through an RFID based cyber physical production monitoring and control in a distributed manufacturing environment in clothing industry the company increased

Fig. 6 Descriptive analytics provide information of past and current events

its production efficiency by 25 %. Furthermore they could spot the bottleneck operation in their production process. By using RFID chips to measure the working time of each operator in the process no computer operator was required to input job tickets. With all the collected data the management levels could easily overview the status of each order and a cloud based production control could plan the incoming orders for the several plants in the distributed environment [15].

Large amounts of data can resolute in such an amount of patterns that it might be impossible to perceive the interactions between the events. In this case, methods from the field of diagnostic analytics are used to analyze the patterns between the events to extract useful information.

3.3 Diagnostic Analytics

In the second step of the data analytics process, the diagnostic analytics, analyzing patterns within the generated data is the major objective. Recurring patterns in collected data can e.g. supply information about repetitive machines sequences in the production process or seasonal variations within a production cycle. This step will be described with the help of a web-based diagnostic tool for production feedback data (WoPS). The tool WoPS was developed and is in use by the department production management of the WZL Aachen University.

WoPS calculates its indicators on the basis of feedback data from an ERP-system. To gain valid key figures, the data should be available for at least two months of production. Additionally, work schedules, resource lists and shift models of the operating system are used as input data.

After uploading that data into the tool, the user is given insight into the quality of his production control and the potential laying in the existing production structure by provision of seven individual indicators. The seven indicators can be structured into three categories: production structure indicators, production control indicators and potential indicators. They allow an overall evaluation of production control and answering questions, why things happened in production as they did.

The first category has high aggregation level and thus provides overall information, not detailed for single machines or orders. The category includes three single indicators: data quality, production process complexity and utilization.

The second category includes bottlenecks, throughput times and work in process (WIP). In the following, the bottleneck analysis is described more briefly as an example, how diagnostic analytics are conducted with the help of the WoPS tool (see Fig. 7 for a screenshot of the tool).

The chart bottlenecks (Fig. 8) shows the process time and the wait time over the given period for each machine. The process time is the sum of lead time and set up time visualized by the grey bars. The blue line represents the wait time of orders at the machines. The comparison of these two factors allows conclusions about the production control performance. Altogether, the machines with long wait times have to be examined critically, to find out the reasons why materials have to wait

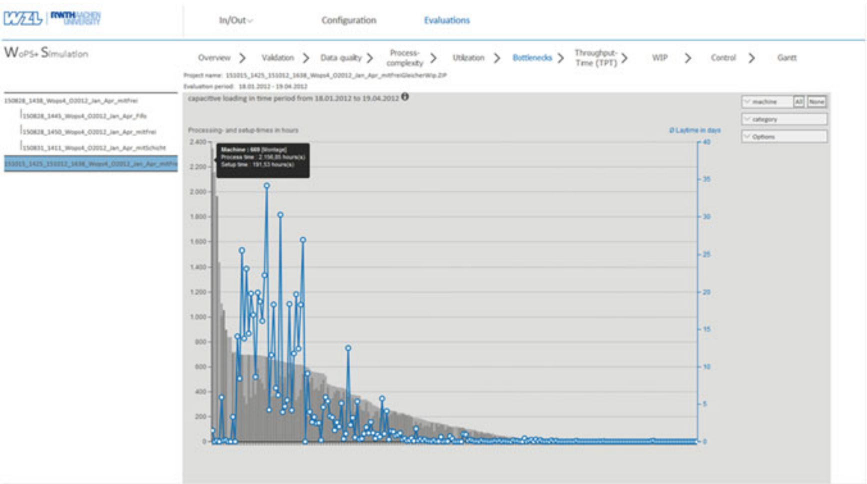


Fig. 7 WoPS screenshot visualizing bottlenecks

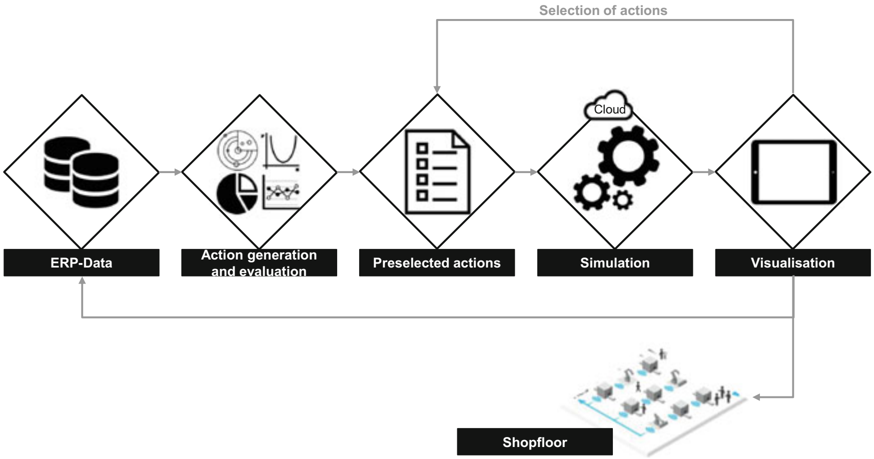


Fig. 8 Descriptive analytics approach in the research project "ProSense"

for handling before these machines. One reason for long wait times could be process related: Materials in small lot sizes are "blocked" (combined to bigger batches) before they enter an expensive machine like an annealing oven, in order to utilize this machine to a bigger extent.

The following four special cases can also be read out of the bottleneck chart in order to answer other questions: The chart enables the user to get the needed information at a glance and to concentrate on the machines with most problems. The described analysis and evaluation is supported by several interactive components in detailed sheets which help the user to concentrate on relevant machines, relevant periods and relevant data. Therefore, a filter for selection of only a part of the given period is implemented. By click and pull, the period can be varied and restricted to relevant periods concerning the future production scenario. Several further possibilities to choose from are implemented. To mention just a few ones, it is possible to choose machines separately or to choose all machines with wait time longer than one day. The option to show machines from one section of the production or to show just wait times or only process times exists as well. With the help of the tool capacity bottlenecks can be identified in respect to machines and staff.

The third category focusses on the potential laying within a given production structure. Therefore, defined changes concerning the configuration of production control tasks are simulated automatically and the results are shown to the user [32].

All in all, the tool WoPS helps conducting diagnostic analytics by researching why things in production happened as they did. Patterns can be identified by analyzing the single charts as well as by combining the various charts of the three categories.

3.4 Predictive Analytics

The third step of the data analytics process, are predictive analytics. Big Commerce companies like Amazon have already implemented predictive analytics in their business. A massive percentage on the Amazon website is assigned to recommend products. Each recommendation is based on predictive algorithms. One enabler for predictive analytics are artificial neural networks, which are used for intelligent data processing. Artificial neural networks are already used these days to predict changes in complex systems. Examples are the prediction of weather phenomenons such as tornados or the prediction of further time series developments like economic growth. Information technology tries to depict the human nervous system by the help of neural networks. The learning ability is hereby produced by the independent activation of connections in the network, changes of weightings as well as by the adding and deleting of neurons [6, 43].

Insufficient or too complicated system support is one shortcoming of todays' IT systems. For the production planner it is hard to find the "right" timing for changes in the production system. Furthermore, it is difficult for him to estimate the interaction between his decisions and the whole production system. Qualified decisions cannot be made due to the lack of tool support.

To solve the problems mentioned above, the production controller needs support: On the basis of real-time executed simulations or model considerations, "what-if"-questions can be analyzed. By comparing different results as well as the probability of their occurrence the most promising strategies are then proposed to the production controller. For such predictive systems, a large number of high-resolution data from the production is the fundamental requirement. Based on this data the artificial neural network learns the coherences of past events and can predict the impact on future decisions.

Artificial neural networks can be applied to different fields of the production control and allow a faster decision making. Striking examples of how high-resolution feedback data and the use artificial neural networks improves the decision-makers' diagnostic capability can be found in the field of pattern recognition, control theory or robotics [8, 25].

3.5 Prescriptive Analytics

Prescriptive analytics the base for generating decision support for many managerial problems and questions. Therefore it enables quantified and reasonable decisions in todays and tomorrows demanding business surrounding. The prescriptive analytics process step will be described with the help of the example of the deviation of decision support for production controllers in the Industrie 4.0 research project "ProSense" [35].

The approach for generating decision support performed in the research project "ProSense" is stated in the Fig. 8.

The use of automatic generated simulation models is a central method in the project "ProSense", which enables drastically shortening the decision-making process in the production control (c.f. Sect. 2.2.3). To provide the needed amount of feedback data, the classical sources of feedback data from production like PDA and MDA systems are expanded by additional sensors, such as RFID sensors, laser and bar code or data matrix scanners. All generated feedback data is cached in the database of the data leading ERP system. A specially developed database application, the so-called Data Collector, saves a history of the feedback data and the past and current planning data in standardized form and thus forms the interface between the ERP system and the simulation software Plant Simulation. On that basis, a simulation model is generated automatically. Data about the existing machines, the production program including all work plans and all necessary master data configure the parameters of the simulation model afterwards. For the parameterization of the control modules for the sequencing, order release and capacity control, the default settings are checked through pattern recognition in the feedback data for plausibility.

To validate the simulation model automatically driven simulation runs are carried out. The results are aggregated based on metrics such as mean and standard deviation of the WIP, throughput time per unit and machine utilization. The aim is to achieve an accurate digital shadow of the real production in the simulation model. Based on the automatically generated simulation model, possible actions are then simulated, leading to an improved capability to achieve the logistic objectives adherence to delivery dates, throughput time, utilization and work in progress. To identify potentially appropriate actions, a list of ten pre-defined generic actions in combination of four pre-defined action areas (machine and personnel utilization, rush orders and order release) is given as input in the simulation run.

Through the automatically generated simulation model that runs on a central server and which offers decision support through the prioritization of actions, the production controller can concentrate on the core tasks without having to worry about the complex data aggregation and analysis. The overall objective is to provide the production controller greater transparency about the actual production status and the effects of different control actions. An increase of the logistical efficiency of manufacturing companies can be reached through the stated approach [35].

3.6 Adjusting Production

Today, most decisions are based on the employees' gut feeling or experience. Based on the results of data analytics, employees are now provided with the information to adjust the production and implement the derived suggestions by the system. By means of the recommendations for the production planner e.g. bottlenecks can be prevented or orders can be shifted due to the lack of parts. Providing these information, new forms of assistance system (e.g. tablets or virtual reality support systems) can be used to give employees guidance how to deal with specific situations.

The decision support system should include probabilities and simulation based real time "what if"-scenarios. Decision proposals for the production planner made by the cyber-physical support system should be made in consideration of the entire production system. This prevents that a change in the schedule in one section of the manufacturing process causes a massive loss in logistical performance in other sections.

4 Summary and Outlook

Globally operating companies are still facing major challenges. The time to react to market requirements is decreasing which is reflected in shorter delivery times. In order to carry out a stable production in this environment, production control must constantly meet new and existing challenges. As a vision of production control, data analytics methods have to be implemented to meet these challenges. In everyday life such decision aids are used already today. They can be found as one example in the rear view camera of modern cars. The system predicts the direction of travel of the car and displays it in the camera image, depending on the chosen steering wheel angle. In this chapter the data analytics process has been introduced and adapted to the production control purposes. In the first step, the *Descriptive Analytics,* effective forecasting involves at first a high quality data basis. In order to control and optimize a production it is extremely important to obtain a digital shadow of the production by the use of sensor technologies. The data quality as an enabler requires some essential factors for the success of forecasting and learning skills. Real-time information in the form of consistent inventory reports must also be made possible on the basis of the part geometry. In the second step, the *Diagnostic Analytics,* the data basis allows a systematic review of the taken decisions. This allows a better visualization of the interdependencies and taken decisions in the past ("What happened?"). To improve the production control a further step is needed to answer the question "Why did it happen". By means of pattern recognition the system allows to reveal interdependencies between the collected data. In the third step, the *Predictive Analytics,* involves the prediction of future conditions of the production environment. A visualization in the form of "weather forecast of production" with a valid corridor from the beginning several hours up to several days later should in this case be the target image. The last step of the data analytics approach is the *Prescriptive Analytics*. Based on the forecast and current state of the production the system provides decision support to gain competitive advantages.

In this chapter the concept of a *"cyber physical production control" (CPPC)* has been introduced to enable companies handling today's and future challenges. The general concept has been discussed regarding the elements data quality, descriptive analytics, diagnostic analytics, predictive analytics and prescriptive analytics. While the concept provides a general description of a *cyber physical production control* further research is needed to substantiate the presented solution principles for the different tasks in order to verify and validate the model's ability.

Acknowledgments The authors would like to thank the German Research Foundation DFG for the kind support within the Cluster of Excellence "Integrative Production Technology for High-Wage Countries".

References

1. Albright SC, Winston WL, Zappe CJ (2011) Data analysis and decision making, 4th edn. South-Western/Cengage Learning, Mason
2. Arbeitskreis Smart Service Welt/acatech (2015) Smart Service Welt. Umsetzungsempfehlungen für das Zukunftsprojekt Internetbasierte Dienste für die Wirtschaft. Berlin
3. Blue Yonder (2015) About Blue Yonder. http://www.blue-yonder.com/en/company/about.html?show=500&url=technology-review-award
4. Broy M (2010) Cyber-physical systems. Innovation durch software-intensive eingebettete systeme. Springer-Verlag, Berlin
5. Brunk T (2014) Wie Daten wirklich wertvoll werden. In: Computerwoche (10). http://www.wiso-net.de/webcgi?START=A60&DOKV_DB=ZECO&DOKV_NO=CW2014030383044056368312447482&DOKV_HS=0&PP=1. Checked 3 Mar 2014
6. Crone SF (2010) Neuronale Netze zur Prognose und Disposition im Handel. 1. Aufl. Wiesbaden: Gabler (Gabler Research [60])
7. Dünnebacke D (2008) Planung und Bewertung des RFID-Einsatzes am Beispiel der Firma Zentis. Vortragsunterlagen zur EDI-Convention. FIR—Forschungsinstitut für Rationalisierung e. V. Aachen. http://publications.rwth-aachen.de/record/230877

8. Eklund A, Ohlsson H, Andersson M, Rydell J, Ynnerman A, Knutsson H (2009) Using real-time fMRI to control a dynamical system. Honolulu, Hawaii
9. Evans JR (2013) Business analytics. Methods, models, and decisions, 1st edn. Pearson, Boston
10. Eversheim W (2002) Organisation in der Produktionstechnik 3. Arbeitsvorbereitung, 4th edn. Springer, Berlin
11. Geiger T (2014) Ein Blick durch die virtuelle Frontscheibe. http://www.tagesanzeiger.ch/auto/technik/Ein-Blick-durch-die-virtuelle-Frontscheibe/story/19120047. Checked 3 Nov 2015
12. Glowalla P, Sunyaev A (2014) Prozessgetriebenes Datenqualitätsmanagement durch Integration von Datenqualität in bestehende Prozessmodelle. Anwendung von komplexitätsreduzierenden Mustern und ihr Einfluss auf Komplexitätsmetriken. Wirtschaftsinformatik 55(6):435–452. doi:10.1007/s11576-013-0391-1
13. Gómez J (2015) Advanced planning and scheduling-system. In: Gronau N, Becker J, Kurbel K, Sinz E, Suhl, L (eds) Lehrstuhl für Wirtschaftsinformatik und Electronic Government, Universität Potsdam, Potstdam. http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/informationssysteme/crm-scm-und-electronic-business/Supply-Chain-Management/Supply-Chain-Management-System/Advanced-Planning-and-Scheduling-System
14. Günther H-O, Tempelmeier H (2007) Produktion und Logistik, 8th edn. Springer, Berlin
15. Guo ZX, Ngai EWT, Yang C, Liang X (2015) An RFID-based intelligent decision support system architecture for production monitoring and scheduling in a distributed manufacturing environment. Int J Prod Econ 159:16–28. doi:10.1016/j.ijpe.2014.09.004
16. Hauptvogel A (2015) "Bewertung und Gestaltung von cyber-physischer Feinplanung". 1. Aufl. Apprimus Verlag, Aachen (Produktionssystematik, Band 6/2015)
17. Hildebrand K (2004) Datenqualität im Supply Chain Management. Ulm: Bonner Köllen Verlag. In: Informatik 2004 – Informatik verbindet, pp 239–243
18. Hodek S (2009) An approach for modular production from mechanics to decentralized control, realized in the SmartFactoryKL. In: ETFA 2009, 22–25 September 2009, Palma de Mallorca, Spain. IEEE, Piscataway. http://ieeexplore.ieee.org/servlet/opac?punumber=5340902
19. Kapp R (2011) Ein betriebsbegleitendes Fabriksimulationssystem zur durchgängigen Unterstützung der kontinuierlichen Fabrikadaption. Univ., Diss.–Stuttgart. Heimsheim: Jost-Jetter (IPA-IAO-Forschung und Praxis, 504)
20. Kortuem G, Kawsar F, Fitton D, Sundramoorthy V (2010) Smart objects as building blocks for the Internet of things. IEEE Internet Comput 14(1):44–51. doi:10.1109/MIC.2009.143
21. Lee EA (2006) Cyber-physical systems. Are computing foundations adequate? Austin
22. López TS, Ranasinghe DC, Patkai B, McFarlane D (2011) Taxonomy, technology and applications of smart objects. Inf Syst Front 13(2):281–300. doi:10.1007/s10796-009-9218-4
23. Mussbach-Winter U, Wochinger T (2012) Bedeutung der Datenqualität für die Zielerreichung der Produktionsplanung und -steuerung. In: Software markt (Nr.12), p 8
24. Nyhuis P, Wriggers F, Fischer A (2006) Development of a decentralized logistics controlling concept. Washington (207). In: International federation of information processing (IFIP), pp 399–405
25. Ohlsson H, Rydell J, Brun A, Ynnerman A, Roll J, Andersson M, Knutsson H (2008) Enabling bio-feedback using real-time fMRI. In: Proceedings of IEEE conference on decision and control
26. Olson JE (2003) Data quality. The accuracy dimension. San Francisco, Calif. Morgan Kaufmann, Oxford. Elsevier Science (The Morgan Kaufmann Series in Data Management Systems Ser)
27. Reuter C, Brambring F, Hempel T (2015) Approach of information provision for high resolution production control. In: Chiu A, Matias A (eds) Operational excellence towards sustainable development goals (SDG) trough Industry 4.0. Central Book Supply, Inc., Manila
28. Schuh G (2005) Produktkomplexität managen. Strategien—Methoden—Tools. Hanser, München
29. Schuh G, Lödding H, Stich V, Reuter C, Schmidt O, Potente T, Franzkoch B (2011) Wettbewerbsfaktor Produktionstechnik. Aachener Perspektiven; [AWK, Aachener Werkzeugmaschinen-Kolloquium 2011, 26. bis 27. Mai; Tagungsband]. Apprrimus, Aachen

30. Schuh G, Potente T, Hauptvogel A (2013) Cyber-physische Feinplanung. Hochauflösende Produktionssteuerung auf Basis kybernetischer Unterstützungssysteme (103). wt Werkstattstechnik—Online 4:336–339
31. Schuh G, Potente T, Thomas C, Brambring F (2013) Improving data integrity in production control. Procedia CIRP 9:44–48. doi:10.1016/j.procir.2013.06.166
32. Schuh G, Potente T, Thomas C, Luckert M (2014) Interactive analysis and evaluation of production control
33. Schuh G, Potente T, Thomas C, Hauptvogel A (2014) Proceedings of the 47th CIRP conference on manufacturing systems methodology for the evaluation of forecast reliability of production planning systems
34. Schuh G, Reuter C, Brambring F (2014) Improving scheduling accuracy by reducing data inconsistencies in production control. In: Narayanan S (ed) Online proceedings of the 25th annual conference of the production and operations management society reaching new heights. POMS, East Lansing
35. Schuh G, Reuter C, Hauptvogel A, Brambring F, Hempel T, Rapp C, Stimming C (2015) Ergebnisbericht des BMBF-Verbundprojektes ProSense. Hochauflösende Produktionssteuerung auf Basis kybernetischer Unterstützungssysteme und intelligenter Sensorik. Apprimus Verlag
36. Schuh G, Stich V (2011) Produktion am Standort Deutschland. Ergebnisse der Untersuchung 2013. Ausgabe 2013
37. Schuh G, Stich V (2012) Produktionsplanung und -steuerung. Grundlagen der PPS. 4, überarbeitete edn. Springer, Berlin
38. Schuh G, Stich V, Nyhuis P, Ortlinghaus P, Franzkoch B, Tröger K, Hering N, Potente T, Hempel T, Schiemann D (2014) Cyber physical production control. In: Integrative Produktion: Industrie 4.0—Aachener Perspektiven, pp 121–141
39. Selke C (2005) Entwicklung von Methoden zur automatischen Simulationsmodellgenerierung. Techn. Univ., Diss.–München, 2005. München: Utz (Forschungsberichte iwb/Institut für Werkzeugmaschinen und Betriebswissenschaften der Technischen Universität München, 193)
40. St. Galler Tagblatt AG (2010) Intellion vernetzt den Alltag—St.Galler Tagblatt Online. St. Galler Tagblatt AG; Internet Tagblatt Medien. http://www.tagblatt.ch/aktuell/wirtschaft/tb-wi/Intellion-vernetzt-den-Alltag;art149,1663076. Updated 29 Dec 2010, Checked 3 Nov 2015
41. Stadtler H, Kilger C (eds) (2000) Supply chain management and advanced planning. Concepts, models, software and case studies. Springer, Heidelberg
42. Stich V, Blum M (2015) A cybernetic reference model for production systems using the viable system model. In: Umeda S, Nakano M, Mizuyama H, Hibino N, Kiritsis D, von Cieminski G (eds) Advances in production management systems: innovative production management towards sustainable growth. IFIP Advances in Information and Communication Technology, vol 459. Springer International Publishing, Cham, pp 169–176
43. Turner S (2014) Office of policy federal highway administration. Defining and measuring traffic data quality. http://ntl.bts.gov/lib/jpodocs/repts_te/13767.html. Checked 3 Nov 2015
44. Vahrenkamp R (2007) Logistik. Management und Strategien. 6., überarb. und erw. Aufl. München, Wien: Oldenbourg
45. VDI 5600 (2007) Fertigungsmanagementsysteme: Beuth Verlag (VDI 5600), p Blatt 1
46. Wang L, Nee AYC (2009) Collaborative design and planning for digital manufacturing. Springer, London
47. Wang RY, Strong D (1996) Beyond accuracy: what data quality means to data customers. J Manag Inf Syst 12(4):5–34
48. Westkämper E, Zahn E (2009) Wandlungsfähige Produktionsunternehmen. Springer, Berlin
49. wetter.com GmbH (2015) Wetterwecker für iPhone & iPad. Ihre Wetter-App von wetter.com. http://www.wetter.com/apps_und_mehr/wetter_apps/iphone/wetterwecker/. Checked 3 Nov 2015
50. Windt K (2013) Robust manufacturing control. In: Proceedings of the CIRP sponsored conference RoMaC 2012, Bremen, Germany, 18–20th June 2012. Lecture notes in production engineering. Springer, Berlin

51. Würth (2015) iBin Bestände im Blick—Der erste intelligente Behälter. Würth Group. http://www.wurth-industry.be/web/de/wuerthindustrie/cteile_management/kanban/ibin_intelligenterbehaelter/ibin.php
52. Zäh MF, Ostgathe M, Wiesbeck M (2010) Ganzheitliches Datenmodell für kognitive Produktionssysteme. WF Zeitschrift für wirtschaftlichen Fabrikbetrieb 105(4):309–315

A Versatile and Scalable Production Planning and Control System for Small Batch Series

Adrian Böckenkamp, Christoph Mertens, Christian Prasse, Jonas Stenzel and Frank Weichert

1 Introduction

Most of today's companies are faced with the increasing need to operate effectively and efficiently. Since many business parameters are less predictable, the global markets became more dynamic, traditional ways of designing and operating production, and logistics systems are not able to handle these new challenges anymore. These challenges are mainly driven by the globalization of supply chains, shorter product cycles, mass customization, and the rising speed of delivery [12, 26, 30]. All these changes are affecting the company organization and logistics. In this chapter, the automotive industry is considered as the application domain. Nowadays, mass-market vehicles are commonly produced with a continuous flow manufacturing system that consists of a combination of highly efficient production and assembly lines. In a broader sense, all vehicles follow the same designated sequence

A. Böckenkamp · F. Weichert
Department of Computer Science VII,
Technical University of Dortmund,
Otto-Hahn-Str. 16, 44227 Dortmund, Germany
e-mail: adrian.boeckenkamp@tu-dortmund.de

F. Weichert
e-mail: frank.weichert@tu-dortmund.de

C. Mertens · C. Prasse (✉) · J. Stenzel
Fraunhofer Institute for Material Flow and Logistics,
Joseph-von-Fraunhofer-Str. 2-4,
44227 Dortmund, Germany
e-mail: christian.prasse@iml.fraunhofer.de

C. Mertens
e-mail: christoph.mertens@iml.fraunhofer.de

J. Stenzel
e-mail: jonas.stenzel@iml.fraunhofer.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_22

of production steps until they are readily assembled. A centralized production planning and controlling approach enables a smooth production flow through the described system but requires significant investment in IT infrastructure and production equipment. Additionally, the approach is best suited for deterministic customer demands which do not exist at all times and in all markets [2, 31].

New trends within the automotive industry, such as the launch of electric vehicles and various technologies in the field of Industry 4.0 [4], challenge existing production systems. Specifically, these challenges include an increased need for flexibility through volatile selling markets, unforeseeable technological developments (e.g., in the field of electric vehicles), shorter production cycles, and the increasing individualization of products [13, 22, 23, 31].

Following the vision of Industry 4.0 (cf. Sect. 1.1), these challenges could be resolved through the concept of smart factories. These factories are scalable and operate at high flexibility through decentralized control strategies, like *Multi Agent System* (MAS) approaches for material flow and program planning, and the utilization of technologies such as the Internet of Things (IoT) and Cyber-Physical Systems (CPS), refer to [4, 32]. Within the scope of these smart factories, the decentralized program planning approach as well as the application of CPS to assembly tasks have to be explored. An approach towards semi central production planning was developed in the SMART FACE project, cf. Fig. 1. The main objective of this chapter is to provide the potentials, advantages, and open questions of this concept.

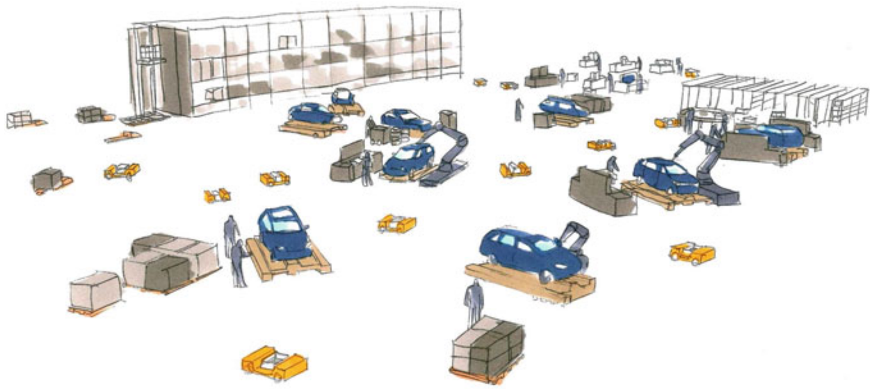


Fig. 1 The vision of SMART FACE: decentrally and autonomously acting units (e.g., assembly stations, automated guided vehicles, industrial robots, etc.) are responsible for transporting and producing goods on a self-organizing shop floor. Human workers are still an important part since they exhibit the largest flexibility that is an important enabler for the concepts

1.1 Industry 4.0 Based on Cyber-Physical Systems

The term “Industry 4.0” has been extensively discussed during the last years, with inconsistent results. In the study by Bauer et al. [3], the authors found 104 different definitions of Industry 4.0. Common to all contributions is the enhancement of production and adding value systems by connecting the real and digital world. Finally, Industry 4.0 leads to a vertical and horizontal conjunction of all process, control, and management levels within and between the companies. Further methods and technologies developed under the general term Industry 4.0 enable an efficient, decentralized and flexible controlled manufacturing of goods or provision of services. From a technical perspective, Industry 4.0 consists of CPS based on embedded systems. They are integrated in a communication infrastructure of an IoT and services.

Following the idea of CPS, several technologies are necessary as an enabler and therefore need a certain maturity level to be applicable. Within the study by Bischoff et al. [5], seven fields of technology have been identified to enable the vision of Industry 4.0 and CPS, respectively. Figure 2 depicts these technologies in detail. For every technology, at least the highest state of development has to be reached in order to achieve the vision of Industry 4.0.

Starting with communication, especially wireless real-time communication is mandatory for CPS applications because usually CPS consist of mobile devices or

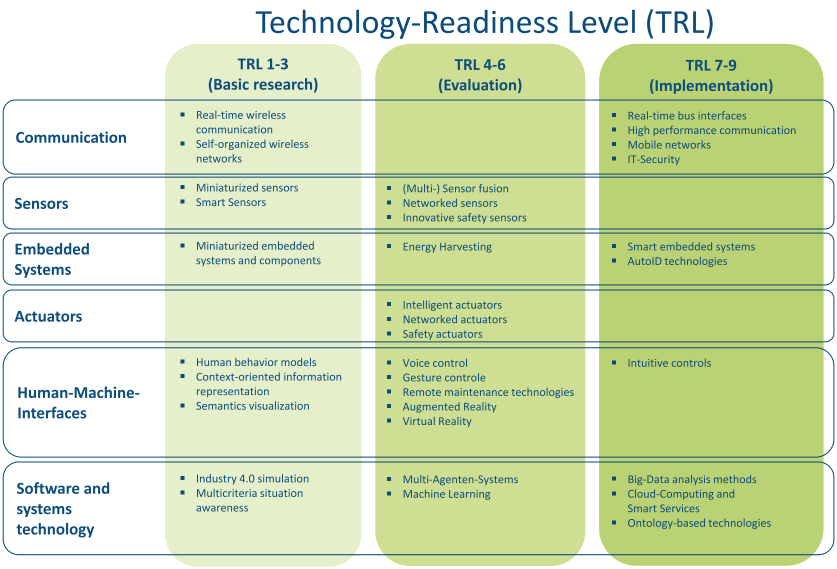


Fig. 2 Mandatory technologies and their current technology readiness level (TRL) to enable the vision of Industry 4.0 based on CPS [5]; at least the highest level of maturity has to be reached for every technology to achieve the vision of Industry 4.0. Current and future research projects should contribute solutions that enhance the maturity level of each single technology

objects that execute local decisions. Sensors provide CPS with the needed data or, in case of smart sensors, with information. Human workers in CPS necessitate seamless interfaces with a high usability. At least efficient software and system technologies like machine learning and MAS ensure a successful implementation of CPS in manufacturing and logistics applications. While contributing in terms of enhancement of technologies like smart sensors, localization, path planning and MAS, the focus within the SMART FACE project is program planning of small manufacturing series.

1.2 Smart Factory Versus Smart Logistics

Within the study by Bischoff et al. [5], German research projects addressing Industry 4.0 topics have been investigated concerning the application field (see Fig. 3) during the last 5 years. The main outcome is a significant number of projects focus on production. Nevertheless, improvement and adoption of production technologies towards self-controlled and autonomous manufacturing systems will not reach the objectives of efficient and self-organized value-adding production networks. Rather horizontal virtual linking-up and even more important physical interconnection of machines in the companies is an essential success factor. Hence, an efficient material flow between production and assembly units as well as between different companies in the value-adding network is a key factor for a successful implementation of Industry 4.0 based on CPS.

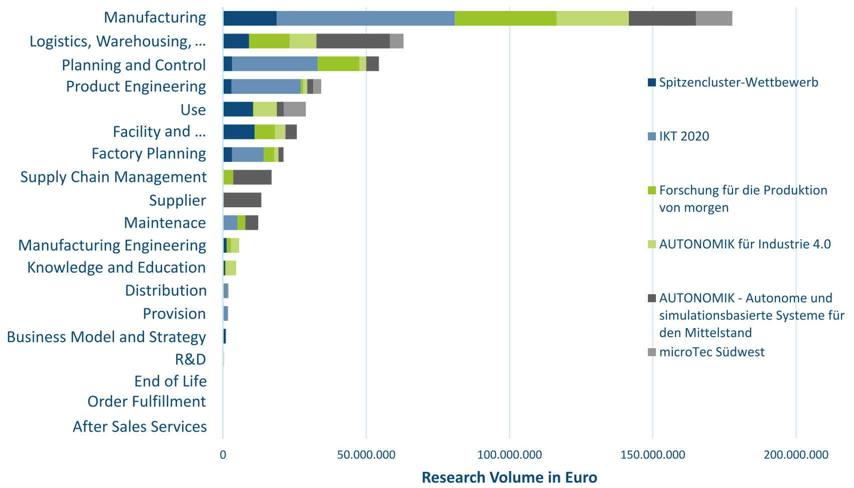


Fig. 3 Research activities related to Industry 4.0 during the last 5 years in Germany [5]. Clearly, most of the projects focus on the manufacturing domain so that upcoming projects should also contribute to other fields of product life cycles. The SMART FACE project enhances the "Planning and control" domain (*third row* of the figure)

By analyzing Fig. 3, it becomes obvious that logistics is an important function within the life-cycle process, but there is still potential to close the gap to production. This is especially necessary for the functions below planning and control. However, the SMART FACE project is contributing to the development of smart and efficient logistic processes and their planning and control. Far above the organizational/informational level, usually addressed in the context of smart factories, the outcome of this project facilitates the integrated implementation of Industry 4.0 based on CPS in practice along the whole supply and production chain.

1.3 Area of Conflict: Deterministic Planning Versus Decentralized Control

Using decentralized control systems in every case causes a strategic decision towards the trade-off concerning the degree of autonomy. Hence, two of these conflicts are described in the following section.

From truth to probability. The first critical point is the trade-off between the determinism (truth) of hierarchical control systems and probability of events in self-organized and autonomous CPS. The use of autonomous and self-controlling entities will raise the level of flexibility because dependencies between different groups of entities can be avoided. Changes in one group do not affect the other parts of the whole system. This also allows a higher adaptability to new requirements or changes in the system performance. Nopper et al. [19] has presented a new life cycle model for flexible (logistic) systems which also includes a model for the benchmarking of flexibility. Unfortunately, the control and behavior of such flexible systems will be no longer deterministic and predictable with current modeling and simulation tools [18]. The behavior of one single autonomous entity is of course deterministic and supports real-time decisions, especially for safety reasons. However, the cooperation or even swarm behaviors of several groups of autonomous entities are no longer deterministic. Decisions depending on the whole systems are taken based on rules and information about the status of the entities, the groups, and even the physical environment. At the end, one prefers (non-deterministically) predicted events with a certain (high) probability to deterministically planned events if the latter will not occur.

From pre-planned sequences to maximum flexibility. The second critical point is the trade-off between clocked planning as well as central control and (time-independent) on-demand behavior of CPS. Within autonomous systems, there is a risk of idle and waiting times which causes a decrease in productivity and have to be avoided. Depending on the individual "skills" of every entity in the system, the mentioned risk decreases with increasing number of skills because one entity does not have to wait for another. Beside an expensive enabling of entities, implementing a local pre-planning in a limited time window at the cost of planning effort could be a another way to solve the problem. Full utilization of the new flexibility demands a change of mind-set, especially with respect to the companies' management. Beside this, management and operators have to learn how to use the new systems. Nevertheless, in case of malfunction or failure in a linked production line, the productivity drops to zero. In contrast to flexible systems as described here, only the disrupted parts are no longer producing. At least the main advantage of on-demand behavior are general changes to the system that can be managed with a minimum effort.

1.4 Overview of Structure

Section 2 gives a short introduction to the conceptual approach behind CPS based and decentralized production systems with corresponding planning functionalities. Therefore three established principles are presented and the research question behind SMART FACE is inferred. Afterwards, Sect. 3 describes the potentials emerging from such a decentralized production scenario and the advantages of such potentials are underlined with examples from the automotive context. The aspect of required planning functionalities is considered in Sect. 4 where a novel approach for the production program planning is introduced—the so called volume cycle. Furthermore, Sect. 5 introduces a concept for a production control system where customer orders are processed in such a way that software agents on the physical instances on the shop floor are enabled to produce products for the customer. A conclusion and an outlook are given in the final Sect. 6.

2 Conceptual Approach

Three well-established basic principles will be presented and adopted to a CPS based production system:

1. "batch planning", a common approach of sequenced planning of manufacturing order fulfillment,
2. "fractal company", a lean production approach consisting of a decentralized structure of self-similar units, and
3. "autonomous working groups", a special variation of division and organization of labor.

An overview is given in Fig. 4.

Compared to continuous fabrication processes of identical goods in a production line, batch production techniques enable the calculation of an optimal number of similar goods which can be manufactured before changing to an other stage of the production process or before changing the product in a production line. Hence, components or goods are produced in groups (batches). The perfect batch size

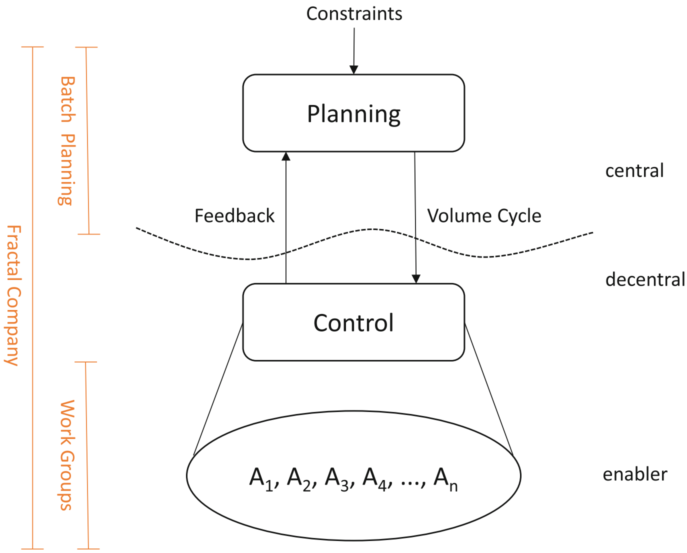


Fig. 4 Conceptual approach of decentralized planning and control based on CPS. The main basic principles can be found on the *left side* of the figure. The concept of fractal companies represents the thematic framework. Batch planning methods help to define the number of orders handed over to working group consisting of CPS and human workers who organize their tasks

(from a line production perspective) is an infinitive lot without changeover stops (and costs). However, an increasing number of goods, variation of products and high dynamics in the markets towards an one piece flow necessitated decreasing batch sizes. A calculation method for economically optimal batch sizes was already developed by Harris [16] and improved by Andler [1]. In the following years and even today, batch sizing and scheduling in production are still recent research topics, especially in the field of operation research (cf., e.g., [14, 15, 25]). A special case of logistics batch sizing can be found in two-step or two-stage order picking systems. In this kind of semi-automated picking systems, a high number of costumer orders are consolidated to picking orders on same-item level in the first stage and manually picked. An automated sorter conveys the single items to the corresponding order-stations in the second step. A main advantage of this type of picking system is the significant reduction of picker's traveling time. This is justified by the fact that a picker can gather all items belonging to a large number of orders with only one stop at the specific bin location instead of executing a new picking cycle for every single customer order. But, well aware that still new orders will arrive in the distribution center, an optimal batch size has to be found to start a picking cycle. A survey on existing methods and a new approach to determine the optimal trade-off between traveling time reduction and utilization of sorter terminals and labor to finalize the orders is given in Ott and ten Hompel [20]. Similar to these general approaches, the volume cycle in the SMART FACE concept is consolidating a certain number of production orders within a defined time slot, adapting several methods of batch sizing presented before. A detailed description of the new production planning approach follows in Sect. 4.

Within the “fractal company” concept [27], rigid interlinking of productions lines dissolves into decentralized self-similar units. A fractal is a graphical formation with a similar appearance independent of up or down scaling the size and a special field of mathematical/natural science chaos theory. First discovered by mathematician Gaston Julia in 1919 (Julia set) and later influenced by Benoît B. Mandelbrot, fractal geometries and its phenomenons were subject of numerous research projects [11]. One result was the fractal basic structure of several processes in nature (e.g., growth of plants, blood cells, etc.). Transferring these fractals to production plants, Warnecke [28] define fractals as autonomous, dynamic and self-similar formations which act as self-organizing and self-optimizing independent company units. Hence, a fractal organized plant consists of self-dependent units with a similar structure. They act autonomously but follow the basic strategies of the overall company. Therefore they need a similar structure as the company itself. Each fractal is controlling and monitoring its continuous improvement and optimization. In case of deviations concerning target goals or limits, the fractal’s activities will be inter-coordinated with the whole company. Hence, an efficient information and communication system is the essential foundation for self-organized fractals. At least this concept is adopted to the thematic bracket between planning (strategic target) and the technical organization of the shop floor (production control) within the SMART FACE approach.

The third approach is to adapt the organization of humanely autonomous working groups to a system of CPS consisting of human workers and automated transport vehicles. Self-organized working teams or (semi) autonomous working groups are a special variation of division and organization of labor. Small teams undertake complex tasks, organization and control of the sub tasks will be done autonomously by the team members. Therefore classical management functions like work preparation, work organization, and qualitative and quantitative output monitoring and control are delegated to the group. Consequently, they are equipped with their own decision and control competence. Depending on the specific type of implementation and tasks to be performed by the working group, different degrees of self-control can be distinguished. Implementing the highest degree of autonomy for the employees, supervisors are no longer necessary because all team members are able to perform every task, hence, hierarchy becomes superfluous. During the post world war II years, Eric L. Twist developed the concept of autonomous working groups mainly motivated by excessive absenteeism due to illness caused by monotone work tasks. At this time, first implementations take place in Japan and only in the seventies of the twenties century in Sweden (Volvo Kalmar). Today, autonomous working groups present the best approach of modern work models. Beside decreasing of illness and increasing of job satisfaction, this concept causes additional beneficial (side) effects, namely high efficiency, high flexibility and reduced planning effort. Nevertheless, it should be noted that there are some disadvantages. At the one hand, it is not a fair system because differences in performance within the group are leveraged by additional work of single team members.

On the other hand, the concept needs a complete change of mind from the management level to every employee. An increasing degree of automation during the last decades have led to a significant reduction of real physical work in industrial environments. Thus, the team members are machine operators, and, as a result, the differences with respect to performance will be leveled and the concept becomes even more attractive today.

The self-similar units described in Warnecke's fractal company are represented by CPSs in this new concept. At the same time, they are–connected with human workers–a new form of autonomous working groups, leveraging the disadvantage of uneven work performance by using a significant number of automated CPS. Autonomous working groups are at least a consequent implementation of the fractal company concept, even if the motivation is unlikely and the idea of working groups is older.

Finally, the research question can be given as follows: How far should autonomy of CPS go or, more precisely, at what point should the strategic planning be changed into a decentralized self-organized control? Main design aspect is the company-individual flexibility within manufacturing and assembly.

3 Exploiting Flexibility Potentials

This section describes the flexibility potentials that emerge when transitioning into a decentralized production concept as described in Sect. 2. Additionally, the advantages of such potentials are explained and illustrated with examples based on the automotive industry.

Typical assembly lines are composed of several process steps which are executed in a strict linear fashion [10]. Even though this has the advantage of optimizing the *fixed* sequence of process steps to maximize the overall performance (e.g., in terms of the number of products per time), a few drawbacks come up, too. Firstly, a single failure in only one of the machines responsible for executing just a single step in the whole sequence causes the entire assembly to halt (breakdown susceptibility). Secondly, with

1. the increase of mass customization (up to lot size 1) by allowing customers to directly configure their desired products on the manufacturer's website, and
2. the continuously emerging market of electrically powered cars (e-cars) the overall product diversity increases significantly. This constitutes a huge challenge for centrally and linearly organized production systems since the enablement of product diversity leads to many different assembly areas which still need to be connected in some linear fashion. However, some assembly areas may not be applicable to all products so that they need to skip specific steps—yet more corner cases. Even worse, such products block machines despite the fact that there is no added value so that they are not available for other products (where value would have been added).

A case study considering the structure and dependencies in the assembly of automotives has previously shown that there is up to 76 % flexibility (cf. [6]). It essentially states that 76 % of the considered assembly steps may be executed before or after the current point in time (during line assembly). In current assembly lines, these execution times are enforced (possibly by executing "empty cycles", i.e., doing nothing). Although a flexibility quantification of 76 % does not provide any hits about the actual flexibility, that is, *how much* a process step can be preponed or delayed, it already indicates a huge unused opportunity.

Considering the general view and context, exploiting the aforementioned flexibility is not easy since new challenges need to be tackled (see Sect. 1.3) and major conceptual changes are necessary (see Sect. 2). As explained, to quantify the expected effectiveness of introducing such concepts, a flexibility graph can be created for the process of interest. Figure 5 shows an exemplary flexibility graph depicting an excerpt from an automotive assembly line, cf. ibid for more details. The vertices (white boxes) represent the process steps of the assembly and the edges represent their dependencies. To process a customer order, it is mandatory to pass *every* node of the graph where each node represents one process step in the whole production process. With respect to the process steps in boxes "A" and "B", the critical path is visualized as a solid line—the longest path through the graph. It is considered as critical because a single delay in one of these processes may result in a delay of the whole product delivery. The interesting part of the graph is given by

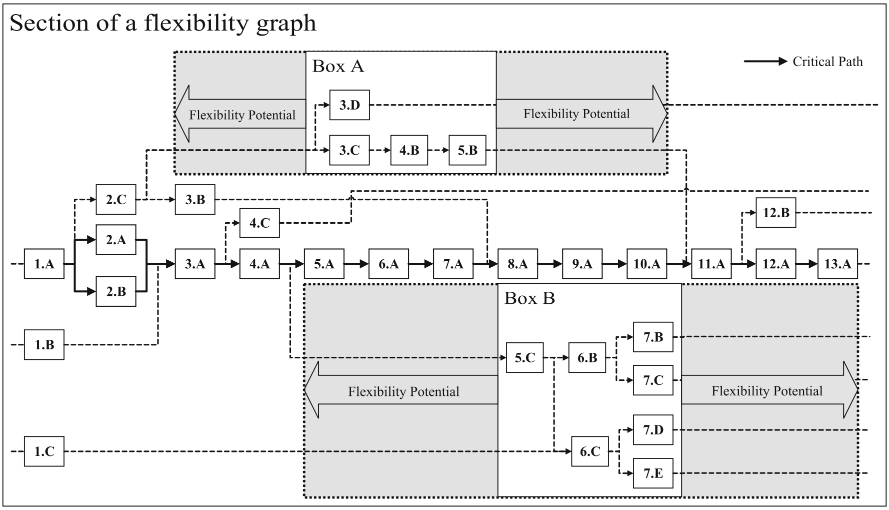


Fig. 5 Exemplary graph showing an excerpt of a production process which demonstrates the flexibility potentials during several steps of an automotive assembly line (from Bochmann et al. [6]). The *white boxes* indicate steps of an automotive assembly process. Their dependencies are given by the edges connecting the boxes. For example, process step "3.A" (e.g., the car battery) requires "2.A" and "2.B" (e.g., the body shell and engine respectively) to be finished beforehand. Note that *all* these steps need to be executed to complete the subpart depicted in this graph

the gray arrows which indicate the flexibility potential for the processed covered in the boxes "A" and "B". In fact, they may be executed earlier ("moving" them to the left) or later ("moving" them to the right).

It remains to describe how the suggested flexibility can be realized and, additionally, how the breakdown susceptibility can be reduced. Two major design aspects are important:

1. *Self-similar assembly islands* (based on ideas of [27, 28]) are composed of multiple, versatile functionality and extensively equipped with (intelligent) sensors to react on environmental changes. This also involves the capability of reacting variably to these changes. For example, assume that a vehicle detects some malfunction nearby an assembly island equipped with an industrial robot. The robot may dynamically provide support by switching to a gripper tool to transfer the load of the faulty vehicle to a replacement vehicle. Clearly, this remains a local decision of the robot; he may reject his support due to some urgent pending assembly task. A notable disadvantage is the requirement of explicit transports between the islands and their integration in the production planning. In contrast to assembly lines, transports are somehow implicitly given in terms of conveyors.
2. *Redundancy in terms of capabilities* provides a way of re-scheduling tasks in case of failure. Examples are multiple assembly islands capable of doing the same tasks, multiple vehicles, etc. This effectively allows to continue production in case of failures and allows one to handle the previously mentioned mass customization. A drawback of this aspect are the resulting costs for redundancy. In practice, one must trade off the investment costs against possible loss of earnings and contractual penalties.

The design also respects scalability: due to the flexible integration of assembly (stations) and transport (vehicles), new industrial buildings may be added to the initial building. Once the vehicles have learned the added areas, more stations and vehicles can be integrated to increase the performance. Furthermore, the design accounts for adaptivity in terms of varying work load: Vehicles may be dynamically added or removed and stations may be switched on and off on demand (saving energy as well).

Finally, it must be noted that human workers are still an essential and irreplaceable element of the smart factory organized in such a decentral manner, cf. Fig. 1. This is justified by the flexibility of humans itself. Because machines are only able to solve tasks and problems which have been considered and technically integrated previously, human workers are always one step ahead in terms of flexibility. However, it can be concluded that the requirements and qualifications of most of the human workers are expected to increase over time. At the same time, workers are being supported by adaptive monitoring and assistance systems to counterpose the increasing machine complexity.

4 Production Planning

One goal of SMART FACE is to streamline the planning process. To understand where the process can be rationalized, it is mandatory to understand the current planning approach which is depicted in Fig. 6 and described in the following. The technological evolution of production systems from embedded systems to CPS affects the planning methodology as well. Due to the fact that the current production system evolves from a rather inflexible environment to a system which adapts to the customer induced variant diversity, there are certain impacts on the planning task. The flexibility graph described in Sect. 3 and the physical changes of production system from a production line to a smart factory affect the planning process on the level of Production Program Planning (PPP) where the planning of a sequence for the customer orders takes place [17, 24]. Figure 6 shows the process of PPP which is a multilevel process. Starting point for the calculations are customer orders and predictions coming from sales and the automotive market. The first planning step is to aggregate the formerly lose orders and to build weekly programs which contain all orders that have to be build within a specific week. The second planning step breaks down the weekly program into a daily program where the orders are distributed over the production week. A goal of this distribution mechanism is to achieve a uniform distribution of the order features. This step is not mandatory and only performed to reduce the complexity for the third and last step which is the sequencing. Here, all orders of a weekly or daily program are brought into a concrete sequence. As it can be seen in Fig. 6, this sequence reminds of a pearl chain.

Due to the fact that the smart factory does not call for a concrete sequence of orders, the planning mechanism can be changed to a batch planning mechanism. The planned batches are called volume cycles. The concept of a *volume cycle*

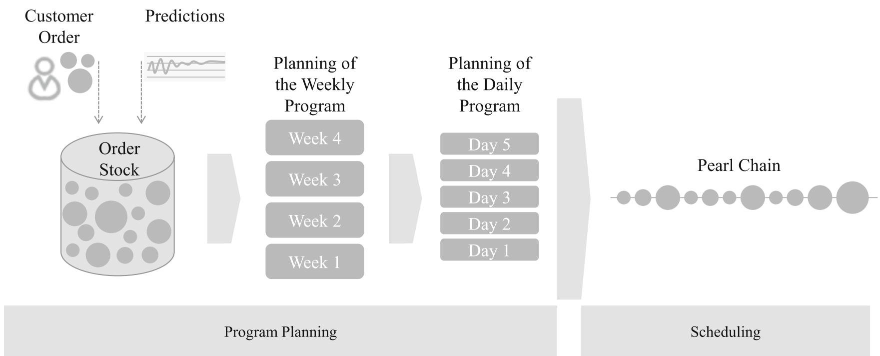


Fig. 6 Present structure of the program planning process and sequencing in automotive industry (cf. [6]). Customer orders and prediction from sales and market are gathered in the order stock. On this basis, the orders are distributed uniformly on weekly programs first, and on daily programs afterwards. The scheduling stage calculates a specific sequence within the weekly or daily programs, the so-called *pearl chain*

bridges the gap between production planning and production control (see Sect. 5) and is the result of the planning mechanism [6].

Compared to the pearl chain as a result of the present program planning methods, the volume cycle does not define a specific order sequence. Instead, it is an unsorted set of orders with varying size and temporal conditions. These temporal conditions are fixed but can vary from volume cycle to volume cycle. As can be seen in Fig. 7, customer orders which can be generated via a product configurator are stored in an order stock. Every order in this order stock is specified by selected properties of the customer product. Amongst others, such properties define the kind of navigation device, capacity of the battery, type of steering wheel or specific driving assistance systems. This information is resolved into the required parts on the one hand and into the process steps of the flexibility graph on the other hand. With the combined information about relevant process steps and required parts, it is possible to determine which assembly islands on the shop floor are appropriate for the production of a specific order. In doing so for every object in a volume cycle, the overall utilization of the production system for a particular set of orders within a certain time frame is shown.

In the planning step of filling the volume cycles with customer orders from the order pool, information about CPS utilization can be used to observe the capacity

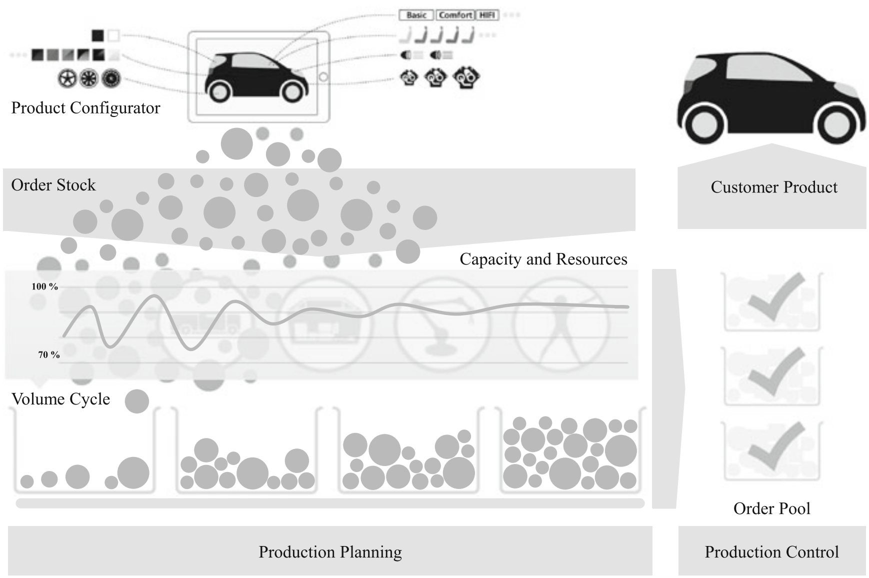


Fig. 7 Production program planning and production control within the SMART FACE project: Customer orders are generated via product configurators and stored in the order stock. Real time information from the shop floor about capacity and resources are used for the planning process. The results of the planing process are filled volume cycles which lead to a balanced utilization of capacity and resources of the production system. These volume cycles are released to the production control so that customer orders can be produced

threshold of the production system as well as the available resources. Information about the current capacity and resources are communicated from the production control to the planning side where sensors of the CPS observe the current situation on the shop floor. Therefore the planning can determine a well-balanced filling of the volume cycles.

Besides the capacity and available resources from the production control side, the production planning can take different target figures into account. These target figures have strong interdependencies, e.g., by maximizing the performance, the lead time is reduced and the storage costs will rise. The following list gives an overview of the target figures considered in SMART FACE:

1. *Maximization of performance:* Under this criteria, the planning mechanism must maximize the number of orders per volume cycle for a given period.
2. *Minimization of lead time:* Here, a fixed number of customer orders is selected which must be produced as soon as possible.
3. *Minimization of use of resources:* The objective is to reduce the number of used resources to produce a given number of orders in a defined period.
4. *Minimization of storage costs:* To minimize the storage costs, the planning mechanism must reduce the number of orders in a volume cycle to a minimum without violating temporal restrictions.
5. *Minimization of mounting costs:* The planning mechanism must compile volume cycles which aggregate orders in a way that reduces the necessity to rearrange and mount the assembly islands.
6. *Adherence to delivery dates:* To match this target, the priority of orders must be determined by their temporal restrictions and by assuring that the distribution of finished goods takes place as soon as possible after the end of production.

Due to the fact that the smart factory is a non deterministic system, the prediction of an exact assembly sequence is not possible. Therefore the Just In Sequence (JIS) concept is not applicable anymore. Due to the planning of volume cycles and the reduction of stocks, the Just In Time (JIT) concept is still relevant for SMART FACE. When the planning side releases a volume cycle, it has to be assured that all the required resources are available so the production control side can choose any order from the released volume cycles.

5 Production Control

When a specific volume cycle (that is, a set of orders/tasks for the cyber-physical units on the shop floor) has been created, it is forwarded to the production control system. This is mainly composed of the MAS which determines which entity (or agent) processes a specific material flow or production order.

A MAS consists of a set of intelligent agents within an environment. Intelligent means that an agent is executing his task optimal under the constraints of his limited information, the environment and his skills. Software agents are programs which are able to act completely autonomous and have total control about their behaviors. Software agents are designed to reach specific goals or to execute specific tasks (cf. [29]).

Within SMART FACE, the CPS of the shop floor are represented by multiple intelligent agents which fulfill specific functions (see Fig. 8). First, the data of the volume cycle is transmitted to a (centralized) repository where it is stored in a data-centric manner. Basic functionality of a Warehouse Management System (WMS) and a Manufacturing Execution System (MES) is realized as an application in order to convert the volume cycle into an agent- and material flow-based representation. The repository communicates with the manager agents within the MAS that are responsible for the coordination of material flow activities (order picking and transportation of goods) and production control activities. These manager agents acquire and manage knowledge about the availabilities and basic properties (or skills) of the other agents, e.g., the skills and availabilities of different types of AGVs or assembly machines.

In general, there are multiple variants the MAS can assign tasks to agents. However, this always involves an (online) optimization problem, thus the manager agents (shown in Fig. 8) enable an optimized scheduling and dispatching of these tasks, e.g., by optimizing the performance of transport tasks using an AGV fleet. As the number of idle vehicles at a specific point in time is known to the transport order manager, this agent needs to solve the task assignment problem based on the knowledge given when triggered. For example, a simplistic solution for the task assignment problem of material transport orders is the evaluation of a score function $s(v_i)$ for each idle vehicle v_i [8]. The score $s(v_i)$ yields a rating for v_i taking into account its current battery level (the higher the better) and its current Euclidean distance to the storage location (the lower the better). A better solution involves, for

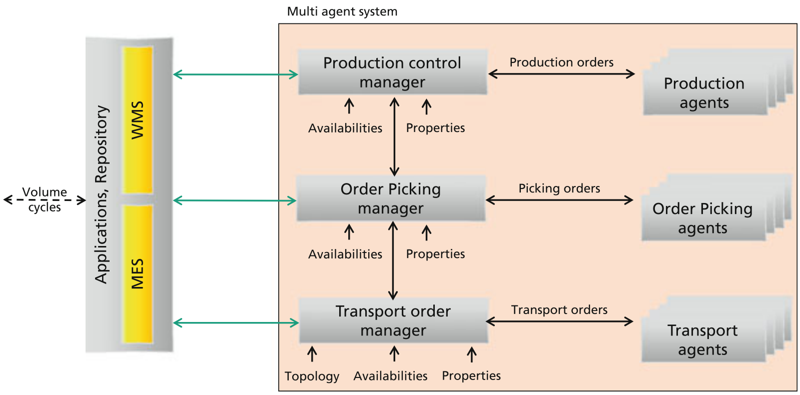


Fig. 8 Simplified overview of intelligent agents representing the CPS in the SMART FACE project. The data of the volume cycle is centrally stored in a repository. Basic functionalities like a Warehouse Management System and a Manufacturing Execution System are operating as applications on this repository. Software agents for order picking, transportation and production control are interacting with the repository

Fig. 9 Initial prototype of production planning and control concepts presented at the Hanover Fair 2015. Within the SMART FACE project a simple, miniaturized and very abstract model of an industrial cyber-physical production system has been implemented. Two miniaturized AGVs representing the assembly order (car models) and two AGVs showing the parts supply (visualized by tablets showing the delivered parts)

example, machine-learning algorithms to learn patterns of when vehicles become idle. This way, the MAS can decide to delay the assignment of a task in favor of selecting an agent which will become available shortly.

Alternatively, the AGVs can be modeled as a set of transport agents and the order picking activities (loading/unloading of AGVs, for an example, see [7]) as a set of order picking agents. Based on the definition of an optimality criterion, e.g., by defining a fuzzy inference system, the transport agents decide which transport order they accept next (cf. [9]). In general, optimality criteria have to match the constraints of the volume cycle (e.g., the latest possible finishing time) and they have to be known by the agents and/or manager agents.

Within the SMART FACE project a simple, miniaturized and very abstract model of such an industrial cyber-physical production system has been implemented (cf. Fig. 9). For this simplified scenario, it has been shown that a combined production planning and control system with decentralized material flow units is feasible. However, now it has to be shown that this system design is transferable to a more complex scenario where production and material flow processes are borrowed from a real automobile production process. This should be done within another demonstrator scenario that will be showcased at Fraunhofer IML at the end of the project.

6 Conclusion and Outlook

The presented chapter is attached to the superordinate questions dealt with in the project SMART FACE. The first question of relevance to the production of small batch series is the trade-off between central production planning and decentralized decisions on the shop floor and the flexibility involved. The volume cycle is an enabler for flexibility in the smart factory, too. Because there is no need to schedule customer orders in a fixed sequence any longer, the flexibility on the shop floor enabled by CPS is fully exploited in the planning phase and at the same time reduces the planning complexity and effort. One aspect of complexity reduction is the elimination of JIS planning methods which are no longer regarded. The task of finding a concrete sequence is not eliminated as a whole but furthermore relocated in the shop floor and therefore on the production control side where the CPS and other agents of the MAS take over the task of sequencing. This does not lead to a less complex system but makes the complexity manageable. In doing so, the planning of volume cycles is just an evolutionary step towards Industry 4.0 where the fixation of a set of customer orders is mandatory to coordinate the supply side. With increasing implementation of Industry 4.0 technologies over the whole supply chain, the planning rather aims at a load-oriented order release with even smaller planning effort.

The second question addresses the concrete form of how CPS will be implemented in Industry 4.0. Within the initial demonstrator, feasibility could be proved for material supply of assembling station organized by a MAS and executed by AGVs. For a further clarification of these issues, an extension of the initial demonstrator will be realized. This system demonstrator allows to represent the interaction between production planning and control system. Furthermore, it will be evaluated if it is even more beneficial for flexibility of modern manufacturing systems when the production planning and the control system share information. The collaborative interaction of actors (e.g., robots, vehicles, etc.), sensor networks, and the communication of these system components will be examined, too. Stepping towards a highly flexible system, the role of human workers as the prototype of an omnipotent allrounder becomes important. Albeit suitable for many applications and tasks human behavior is also difficult to predict which constitutes a challenge for man and machine cooperation. For future work, several requirements for the integration of humans in CPS will be evaluated (cf. [21]).

Acknowledgments This work was supported by the German Federal Ministry for Economic Affairs and Energy (BMWi) under the "AUTONOMIK für Industrie 4.0" research program within the project SMART FACE (Grant no. 01MA13007). The stakeholders of this project consist of industrial companies and research institutions, namely Logata Digital Solutions, F/L/S Fuzzy Logik Systeme, Lanfer Automation, Continental, SICK AG, Volkswagen AG, TU Dortmund University, and Fraunhofer IML.

References

1. Andler K (1929) Rationalisierung der Fabrikation und optimale Losgröße. Ph.D. thesis, University of Heidelberg
2. Aurich JC (2014) Automobilproduktion. Springer Vieweg
3. Bauer W, Schlund S, Marrenbach D, Ganschar O (2014) Industrie 4.0—Volkswirtschaftliches Potenzial für Deutschland. BITKOM
4. Bauernhansl T, ten Hompel M, Vogel-Heuser B (eds) (2014) Industrie 4.0 in Produktion, Automatisierung und Logistik. Springer Vieweg, Wiesbaden

5. Bischoff J, Hegmanns T, Prasse C, ten Hompel M, Henke M, Wolter D, Braun N, Guth M (2015) Erschließen der Potenziale der Anwendung von Industrie 4.0 im Mittelstand. In: Studie im Auftrag des BMWi
6. Bochmann L, Gehrke L, Böckenkamp A, Weichert F et al (2015) Towards decentralized production: a novel method to identify flexibility potentials in production sequences based on flexibility graphs. Int J Aut Tech 9(3)
7. Böckenkamp A, Weichert F, Rudall Y, Prasse C (2016) Automatic robot based unloading of goods out of dynamic AGVs within logistic environments. In: Clausen U, Friedrich H, Thaller C, Geiger C (ed) Commercial transport: proceedings of the 2nd interdisciplinary conference on production, logistics and traffic 2015. Lecture notes in logistics. Springer International Publishing, Heidelberg, pp 397–412
8. Böckenkamp A, Weichert F, Stenzel J, Lünsch D (2016) Towards autonomously navigating and cooperating vehicles in cyber-physical production systems. In: Niggemann O, Beyerer J (ed) Proceedings of the 1st conference on machine learning for cyber physical systems and industry 4.0 (ML4CPS), 2015. Technologies for intelligent automation. Springer International Publishing, Lemgo
9. Branisso LB, Kato ERR, Pedrino EC, Morandin O, Tsunaki RH (2013) A multi-agent system using fuzzy logic to increase AGV fleet performance in warehouses. In: 2013 III Brazilian symposium on computing systems engineering (SBESC), pp 137–142
10. Chow WM (1990) Assembly line design: methodology and applications. Manufacturing engineering and materials processing. Taylor Francis, New York
11. Crilly T (2009) Fraktale. German. In: 50 Schlüsselideen Mathematik. Spektrum Akademischer Verlag, pp 100–103
12. De Meyer A (1998) Manufacturing operations in Europe: Where do we go next? Eur Manage J 16(3):262–271
13. Dijk M, Orsato RJ, Kemp R (2013) The emergence of an electric mobility trajectory. In: Energy policy 52.0. Special section: transition pathways to a low carbon economy, pp 135–145
14. Drexl A, Kimms A (1997) Lot sizing and scheduling—survey and extensions. Eur J Oper Res 99(2):221–235
15. Fumero Y, Corsano G, Montagna J (2012) Planning and scheduling of multistage multiproduct batch plants operating under production campaigns (English). Ann Oper Res 199(1):249–268
16. Harris FW (1913) How many parts to make at once. Mag Manage 10(2):135–136
17. Holweg M, Pil FK (2004) The second century. Reconnecting customer and value chain through build-to-order; moving beyond mass and lean production in the auto industry. The MIT Press, Cambridge
18. Hülsmann M, Windt K (eds) (2007) Understanding autonomous cooperation and control in logistics: the impact of autonomy on management, information, communication and material flow. Springer, Berlin
19. Nopper J, Follert G, ten Hompel M (2010) Gains in the lifecycle of adaptable, self-organizing material handling systems. In: International material handling research, vol 11, pp 122–142
20. Ott M, ten Hompel M (2006) Einfluss einer dynamischen Auftragssteuerung nach dem Floating-Batch-Prinzip in zweistufigen Kommissioniersystemen. Logist J
21. Rinkenauer G, Böckenkamp A, Weichert F (2016) Mensch-Roboter-Interaktion im Kontext von Industrie 4.0: Annähern-Meiden-Tendenzen als ein Bewertungsmerkmal für die affektive Qualität der Interaktion? In: 62. GfAFr ühjahrskongress: Arbeit in komplexen Systemen. Digital, vernetzt, human?!
22. Schoensleben P (2009) Changeability of strategic and tactical production concepts. CIRP Ann—Manuf Technol 58(1):383–386
23. Schuh G, Wemhöner N, Friedrich C (2006) Scenario-based lifecycle analysis of manufacturing systems. CIRP: J Manuf Syst 35(2):129–135
24. Schwede C (2015) Integration von Auftragsreihenfolge- und Distributionsplanung in der Automobilindustrie. Dortmund, Praxiswissen

25. Tempelmeier H, Copil K (2015) Capacitated lot sizing with parallel machines, sequence-dependent setups, and a common setup operator (English). OR Spectr 1–29. doi:10.1007/s00291-015-0410-2
26. Wahab IMM, Desheng W, Chi-Guhn L (2008) A generic approach to measuring the machine flexibility of manufacturing systems. Eur J Oper Res 186(1):137–149
27. Warnecke H-J, Hüser M (1995) Lean production. Int J Prod Econ 1(41):37–43
28. Warnecke H-J, Hüser M (1996) Die Fraktale Fabrik-Revolution der Unternehmenskultur. Rowohlt
29. Weiß G (1999) Multiagent systems: a modern approach to distributed artificial intelligence. Intelligent robotics and autonomous agents series. MIT Press, London
30. Wiendahl H-P, ElMaraghy AH, Nyhuis P, Zäh M, Wiendahl H-H, Duffie N, Brieke M (2007) Changeable manufacturing—classification, design and operation. Ann CIRP 56:783–809
31. Zubaryeva A, Thiel C, Barbone E, Mercier A (2012) Assessing factors for the identification of potential lead markets for electrified vehicles in Europe: expert opinion elicitation. Technol Forecast Soc Change 79(9):1622–1637
32. Zuehlke D (2010) Smart factory—towards a factory-of-things. Ann Rev Contr 34(1):129–138

Part VI
Evolution of Workforce and Human-Machine Interaction

CPS and the Worker: Reorientation and Requalification?

Ayad Al-Ani

1 Enter the Process Worker

In his book about the "rationality" of economic reasoning, the French theorist André Gorz referred to an analysis of the "process worker" that had been published by an individual using the pseudonym "Ilnox" in the communist daily "il manifesto". The process worker, explains Gorz ([19], 111ff.) citing Ilnox, is the by-product of computerization, which represents a new interface between the production process and the worker. The skilled and unskilled worker of the production line will cease to have any physical interaction with the product, but will now focus mainly on controlling and maintaining the production process. Thus, computerization will impose its own standardizations, such that the kind of work done, irrespective of its location or even industry (brewing, power plant, pasta production …), will essentially be the same—monitoring and controlling the production process at a distance, via display screens. The effects of this transition for the process workers will be hard to underestimate and somewhat contradictory: first of all, the monitoring and controlling tasks, and the context skills needed to complete them, will be somewhat more complex than the current tasks and skills, so some re-skilling will take place. Secondly, these skills will be accessed more easily and will be transferable across more or less all industries and locations, giving the process worker more mobility. At the same time the skills in use may be rendered

A. Al-Ani (✉)
Humboldt Institute for Internet and Society, Berlin, Germany
e-mail: ayad.al-ani@hiig.de

A. Al-Ani
Chair of Business Information Systems and Electronic Government,
Potsdam University, Potsdam, Germany

A. Al-Ani
School of Public Leadership, Stellenbosch University,
Stellenbosch, South Africa

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series
in Wireless Technology, DOI 10.1007/978-3-319-42559-7_23

commonplace, because few company or industry characteristics will be relevant, giving the process worker little advantage over others in competitive situations. Finally, the work to be done will be quite dull. Monitoring events might call for action sometimes, but they will mostly require passive observation. The worker will create nothing, but this nothingness will drain him. Interestingly, Gorz did not view this development too negatively: It will at least free the worker from hard, physical labour and give him a chance to counter balance dull activities with a rich private life outside the factory (ibid, 148).

This development and Gorz's conclusion might be surprising to some extent, but it appears to be the climax of a consistent chain of events. For the longest time, machines competed with man on efficiency. Instead of automation being designed in such a way as to use the skills of the worker, man was increasingly degraded and being fitted into mechanized processes ([14], 259, [10]). Finally this battle seems to be over. Men—so it seems, must go elsewhere in search of identity or even humanity, as Moravec ([26], 134ff.) added with a twinkle: labour is such a minimalist goal!

2 The Opening of the Lights-Out Factory

The goal has always been to remove humans from the production process ([27], 333). The course of this path taken had economic but also political reasons:

> The idea of replacing humans with machines was nothing new, of course. From the Great Depression, through the war, and into the Cold War, the nexus of labor and manufacturing technology was a continual source of both innovation and conflict. […] in 1948, there was widespread concern among the U.S. business elite that organized labor had become too powerful: Union membership had soared during the war, as had the number of strikes. Many in government and industry also worried that the United States lacked the industrial might of the Soviets. Clever, electronics-infused, self-guided machinery promised a solution to both concerns. [11]

In some instances we seem to have reached this goal. The most striking example is the lights-out factory: a production plant so completely automated that robots could turn out the lights by themselves (a very unlikely prospect, as these factories can and should work 24/7). The first examples of this kind of production facility are quite impressive:

> The brightly lit single-story automated shaver factory is a modular mega machine composed of more than 128 linked stations-each one a shining transparent cage connected to its siblings by a conveyor, resembling the glass enclosed popcorn makers found in movie theatres. […] Each of the 128 arms has a unique 'end effector', a specialized hand for performing the same operation over and over and over again at two-second intervals. One assembly every two seconds translates into 30 shavers a minute, 1,800 an hour, 1,304,000 a month, and an astounding 15,768,000 a year. ([25], 66f.)

In this factory, the role of humans is minimized and seems to even exceed Gorz's vision of the process worker:

> Meanwhile, a handful of humans flutter around the edges of the shaver manufacturing line. A team of engineers dressed in blue lab coats keeps the system running by feeding it raw material. A special 'tiger team' is on-call around the clock so no robot arm is ever down for more than two hours. Unlike human factory workers, the line never sleeps. (ibid, 67)

The workers in this plant are, obviously, concerned with controlling and monitoring, although it seems this role will also be taken over by machines sooner or later; the robots themselves already know how to sort out mistakes. The system is therefore tolerant of small errors. Thus, not many workers are needed at all in this kind of factory—in our example it was less than 10 (ibid). The lights-out factory, so it seems, has already surpassed the need to interact with humans intensively:

> Many of the new production methods in this next revolution will require fewer people working on the factory floor. Thanks to smarter and more dexterous robots, some lights-out manufacturing is now possible. FANUC, a big Japanese producer of industrial robots, has automated some of its production lines to the point where they can run unsupervised for several weeks. Many other factories use processes such as laser cutting and injection moulding that operate without any human intervention. And additive manufacturing machines can be left alone to print day and night. [35]

3 What Skills Now?

As the role of the workers in CPS-dominated production is rather questionable, if not marginal, it should come as little surprise that views about the kind of workforce and skills needed, are rather vague ([30], 9ff.). Consequently, and rather in line with Gorz's vision of the process worker, the idea that humans are mainly there to fix errors and to interfere with the CPS if distortions occur seems to be the common ground ([18], 527, [31], 9f.). In a rather comprehensive view on qualifications for CPS-dominated work processes, Pfeiffer [30] is quick to explain that not only monitoring skills are needed but that the nature of the process workers requires a huge amount of "working capacity", that is skills that are needed in order to fix things when needed. This is also in line with Gorz, who conceded that the process worker seems to have long phases of monotonous observations intermingled with hectic phases of problem management. Recalling the picture of the lights out-factory, we should now suspect, however, that this error fixing should pretty quickly surpass the generic skills of the process worker, as detailed and sophisticated knowledge of the production process and the robots would be required. However, Pfeifer, who develops a comprehensive overview of the required skills of the worker in CPS-intensive processes, reaches the conclusion, that the majority of new skills will probably lie in the area of data management, data privacy and data security (ibid, 26ff.). In contrast to Gorz, she furthermore assumes that the company specific characteristics she describes as the "offline side" of the CPS production will become more important: Anything that cannot be fitted into given data structures but that remains important for the production process (ibid, 33); although at this stage, we are left to wonder what this company specific knowledge might be.

4 The (Temporary?) Return of the Gods

In a provocative claim, the science fiction writer and computer scientist Vernor Vinge put forward the notion of a computing singularity in which machine intelligence will make such rapid progress that it will cross a threshold and then, in some yet unspecified leap, become super human ([25], 9pp.]. If Singularitarians are right, this transformation will lead to human labour becoming surplus: There will be fewer places for human beings in the resulting firms and economy. This has certainly not happened yet [22]. A remarkable company policy shift that suggests that there are limits to automation was the recent decision of Toyota to systematically re-integrate humans back into the production process.

> After pushing its automation processes towards lights-out manufacturing, the company realized that automated factories do not improve themselves. ([25], 90]

The return of the extraordinary craftsmen, known as *Kami-sama*, or gods, who, in the traditional company, had the ability to "do anything" with a focus on improving the production process, points to another important role for the worker: not only supervising the automated production process but also serving as a kind of "role model" for the robots and production lines:

> These gods […] are making a comeback at Toyota, the company that long set the pace for manufacturing prowess in the auto industry and beyond. Toyota's next step forward is counter-intuitive in an age of automation: Humans are taking the place of machines in plants across Japan so workers can develop new skills and figure out ways to improve production lines and the car-building process. [37]

From this point of view, the robots and production systems must learn from humans, requiring refined and extraordinary workers who possess deep skills to be recreated in machines:

> 'We cannot simply depend on the machines that only repeat the same task over and over again,' […] 'To be the master of the machine, you have to have the knowledge and the skills to teach the machine'. (ibid)

Singularitarians, of course, would argue that this is a mere interim partnership between humans and robots, during which human knowledge is transferred, and at some point, creativity will be transferred too or will even arise on its own in some brilliant machine of the future ([25], 90). After all, the self-learning machine was the starting point of Silicon Valley [11].

For now, however, there is quite an intriguing point to make: The process worker model falls short in describing the role of humans in modelling robots and CPS assembly lines ([17], 54). The skills needed here are not only the average skills, those easy to access and to transfer, but also deep skills in all the moves and steps needed to assemble the product. Here we find the overlap between the off- and online world of CPS production: Companies that manage to re-create sophisticated human skill levels in their machines and robots will become superior to others that are less capable, as is the case in IT programmes that are better than others, although the same technology and programming language is in use.

5 The New Factory: Connecting the Dots

The often vague and general character of some views on the skills of the factory worker might have to do with the fact that observers sometimes implicitly assume that the factory will continue to exist in its present state in the future. There is, however, little reason to believe that production lines will change dramatically while leaving the factory structure as it is. Sometimes, scholars perhaps overlook that in linking up and interconnecting machines, robots, supply stores and customers, CPS will create opportunities for "opening up" the value chain of the factory to actors outside the traditional boundaries of the organisation [32, 34]. The significance of this development has to be understood against the backdrop of a hitherto de facto "closure" of factories:

> […] to make their just-in-time manufacturing work, Toyota limits the number of suppliers that it deals with and tightly integrates its operations with these suppliers […] In other words, Toyota has been able to achieve high flexibility in its operations by closing its system and limiting the diversity of participants. ([20], 81f.)

Now, CPS will be key in enabling totally new configurations of producers, suppliers and customers. This new setting has been labelled "Open Manufacturing" and should be considered the organisational twin of CPS (see Fig. 1).

If the experience of the well advanced and already considerable "open" IT industry is any indication, than the product of the Open Factory will be a combination of publicly available parts (commons) and proprietary enhancements [38]. This mix will cause a massive disruption for the production line worker: the openness of the product and the factory will be determined by the flexibility of the production line and will ultimately constitute the company's comparative advantage and value creation ([12], 21ff.).[1]

In this new mode of production, different producers work on products that are open in the sense that their patent is public (Tesla) or that standardized and open interfaces to the product (API: Application Programming Interfaces) are available and developers are invited to participate in the development of applications (Watson Cloud). Thus, even developers outside the current boundaries of the factory have access to the product blueprint and participate in its refinement and ongoing development (hacking). These developers sometimes labelled Tinkerers can also collaborate with each other using open working spaces (Fab Labs) that are connected and drawn to the Open Factory to elaborate their ideas, evaluate and improve the

[1]As it has been pointed out, the value of CPS does not rest in the mere connection of different elements of production. In addition, the CPS enabled assemblage will create additional functionalities (and value).

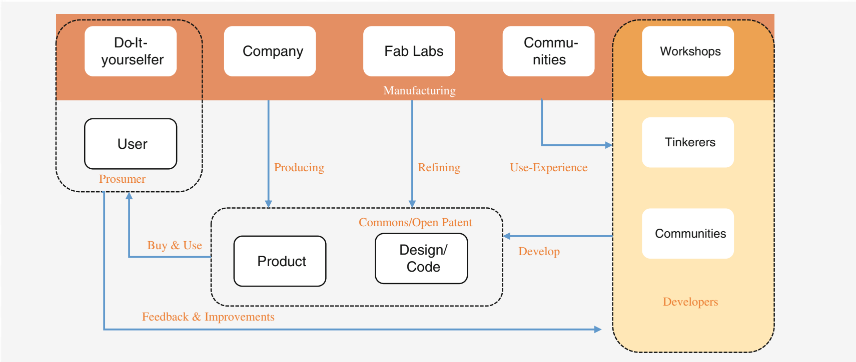


Fig. 1 Open manufacturing. *Source* ([7], 195)

designs of others and participate in the production process ([6], 59ff., [33, 36]). The tested designs would then be sent to the production lines to be "printed out", if the series is large enough or being replicated in a decentral manner using 3D printers.

It is obvious that working as a production "worker" in this kind of environment would be quite a different task than in the lights-out factory. Not so much because the physical production is different in principle, but because this production needs to be connected and reconfigured to cope with different kinds of design inputs and manufacturing options [2]. The production worker in this setting, thus, will embrace more "design thinking instead of production thinking" ([9], 12). In this context, some of the typical digital buzz will make more sense:

> A lot of collaborative and cross-cultural competencies will be required to be able to work in network environments sustainably. On the technical side: connecting the network will mean a lot of standardization. Therefore, the technical competency profile will be rather T-shaped and interdisciplinary than specialized. Analytics specialists will have to work across business models, production processes, machine technology and data-related procedures. (ibid, 13)

By introducing the Open Factory concept into the analysis, a vast array of new skills and tasks now comes into consideration (see Fig. 2).

The "worker" in this setting will need knowledge about the elements of the CPS enabled production that need to be reconfigured: adequate machinery (all-purpose open source MultiMachines, prototyping …), on demand infrastructures (online 3D printing services, Design for download …), Internet of Things (trackable objects, Sensor Commons …), Open Development methods (modularization, open development, crowdstorms …) and social movements (Maker Movement, Open Source Development, Hardware Hacker …). Thus, the worker

> […] will set up determining factors of production, design, install and maintain complex cyber-physical systems and define the rules of production […]. ([32], 7)

The knowledge about those elements will enable him to constitute the "factory" according to the specific requirements of the product and marketing strategies: The "factory worker" evolves into an enabler or architect of the open factory.

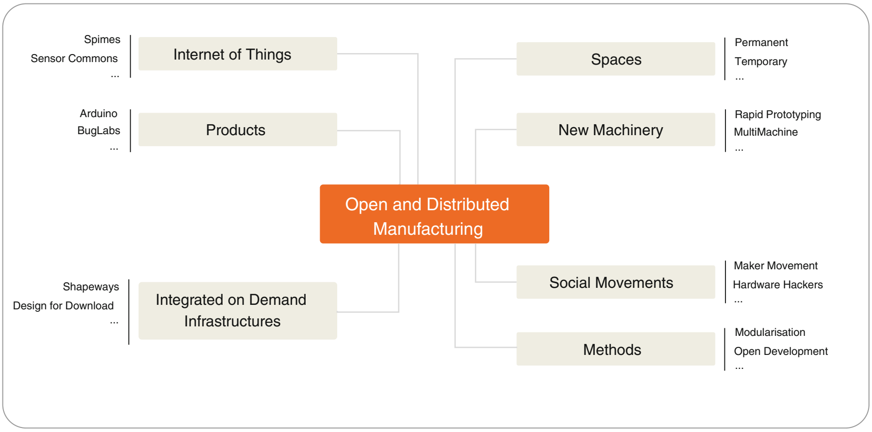


Fig. 2 Skills and tools of open manufacturing (excerpt). *Source* ([7], 210)

6 Worker and Management: Converging Roles?

The new tasks of the "worker" that have been suggested here, obviously have little to do with the original concept of the factory worker. But the changes do not stop here. If we extend Gorz's argument, we can anticipate, that via the influence of ubiquitous social media, automatization and the opening of the work structures enabled by CPS, a kind of convergence of skills, tasks, roles and ultimately profiles in the workplace will take place. Thus, not only will the tasks and skills of "workers" be related *between* different industries, but the many roles *within* the factory may to a relevant extent assimilate as well, at least at the task and skill level. If work will be done mainly using virtual interfaces, (organising production lines, organising customer communities, communicating to employees …) than we can reasonably expect rather similar skills in the new factory across many roles and levels, perhaps even to the extent, that the traditional schism between worker and manager will become less important or even obsolete ([3], L2f.). At least—if we do not wish to engage in the discussion of converging roles and levels we can concede that workers will become more "empowered":

> In a more interlinked world the function of employees will shift away from simple operators towards decision-makers that are actively involved in the decision making process, which focuses not on selective optimization but also considers the overall contexts. ([32], 7)

Of course, there will be specializations, in the sense of roles having different focal points (organising employees, customers or production lines) and deep unique expertise at various points, but the tools used and the resulting profile requirements will be analogous: the digitization of work will lead to an abstraction of different functions and tasks, with the consequences that they will be quite similar although they will have different ends (i.e. addressing customers and connecting producers via the same technology).[2]

At this point, we can only vaguely speculate on what this convergence would mean from the perspective of an economy based on division of labour and specialization. We need to consider that automation hitherto was mainly being used as to reinforce the role of management.

> Automation was designed through the state system to demean and degrade people to de-skill workers and increase managerial control. ([14], 259)

History reminds us, that existing power structures will not be overridden by mere technological advancement or inexorable market forces ([24], 63ff., [28], 3ff.) But structures and institutions are rather shaped according to the political will of the involved actors and parties that—paradoxically often attempt to rationalize their strategies ex post as the best economical or technical solution: might makes right ([29], 151ff.). Although the outcome of this transformation process is therefore highly uncertain, as small actions by groups here and there may shift the vectors and the institutional forms in radically different directions, there are already indications that stakeholders are recognizing this alteration of the power bases, and the first realignments and regroupings seem to be taking place ([4], 11ff., [39], 132).

7 Conclusion: Moving up the Ladder

Starting from the first perspectives of the 1980s, covering current developments and also anticipating the factory of the future, there are three possible models for the human in the CPS production line:

- *The process worker*: This worker is mainly concerned with monitoring the more or less fully automated production process. This rather limited role is the end point of a development that wishes to eliminate humans from the production;
- *The role model*: This role brings back the human as a template for mechanical skills that need to be reconstituted in machines. The human here becomes an extraordinarily skilled and versatile worker, who works in enclaves (workshops) to retain and refine skills that are not used in the production process directly, but will be transferred to the machines in order to gain the upper hand in the configuration of the production process and to maintain competitive advantages;
- *The architect*: Here the factory worker becomes the enabler and configurer of the CPS production, connecting different developers, production lines and customers.

Clearly, these roles require very different sets of skills. Perhaps it is useful to imagine this reorientation as being enabled by a (re-)qualifying learning path for the

[2]Gorz would of course have used the term “alienation” instead of “abstraction”.

present worker. Staying in the inferior role is not an option. Not only is this work quite unchallenging, there will be little demand for it in the future:

> Fortune will instead favor […] those who can innovate and create new products, services, and business models. [13]

When looking at the three models as a kind of development path, however, it is possible to imagine that process workers will evolve to (re-)discover their abilities and become masters of certain skills sets that will, in turn, lead to robots and production lines that are built to their abilities and likings. From this position, it is not a huge step to become the architect of the entire factory layout, which will be in constant flux in order to connect the different type of developers, customers (acting as producers) and the production lines that need to serve these groups. This movement, in fact, has been with us for some time:

> It is sometimes said, that the mass manufacturing ended the role of the craftsman, who was replaced by heavy machines stamping out identical parts. In fact, craftsmanship moved into the engineering department, where products are designed and their means of manufacture devised. ([26], 128)

Learning the skills needed for this qualification—one may argue—will be difficult and not possible for the worker. But Digitization and its new forms of organisation has always been about enabling skills that were neglected and suppressed by the traditional hierarchy ([1], 121ff., [8], 29ff.). Thus, we can anticipate that digitally augmented learning paths will be available to anyone and may make ascending the ladder easier and more feasible [5]. The worker would ultimately behave like an "Edupunk" somebody who could create his learning path not only by using elements of traditional education but also from downloaded learning content and peer learning:

> The way I look at it, a complete personal learning plan ought to have four parts: finding a goal and the credentials or skills needed, formal study, experiential education, and building a personal learning network. ([21], 137)

This educational task will be become an important part of the workers role, and will mainly be his responsibility, if the experiences of freelancers who already operate in a less binding and more open working relationships offer any indication:

> We not only take all the risks of our job moves, we assume the task of taking care of our creativity of investing in it, and nurturing it. […] Increasingly workers have come to accept that they are completely on their own-that the traditional sources of security and entitlement no longer exist, or even matter. ([16], 99)

We can furthermore suspect that companies will develop demographic HR-strategies that will shape and influence the development path described here:

> The VW board member for human resources, said robots would fill some of the retiring baby boomers' jobs, not people. He did however insist that robots would take over the more monotonous or unergonomic tasks allowing human employees to focus on more highly skills jobs. 'We have the possibility to replace people with robots and nevertheless we can

continue to hire the same amount of young employees. Or put the other way: we would not be able to compensate for this outflow of retirees by [hiring] young employees'. [23]

The worker—so it seems—must embrace some kind of *learnership*, the ability to be a competent author of *"[…] one's individual learning journey by leveraging the vast learning opportunities offered by personal networks and the global marketplace*" ([15], 265).

By introducing the need to model CPS according to human skills and recognizing the changing nature of the factory, we suddenly have a wide and rich array of different tasks and profiles for the human, which have somehow been obscured by the image of the process worker. By focusing on the process worker image exclusively, we may have followed a practice that recognized the factory worker as a steady and fixed "factor" of the production plant, unable to fully use his or her capacities and take any active part in the development and refinement of his or her skills. This perception—one must hope—can now be left behind, just like the old factory.

References

1. Al-Ani A (2013) Widerstand in Organisationen. Organisationen im Widerstand. Virtuelle Plattformen, Edupunks und der nachfolgende Staat. Springer, Heidelberg
2. Al-Ani A (2015) Arbeiter zwischen Knechtschaft und totaler Befreiung. Zeitonline. http://www.zeit.de/karriere/2015-05/digitalisierung-arbeiter-fabriken. Accessed 12 Dec 2015
3. Al-Ani A (2015) Wie Globalisierung, Technisierung und der Algorithmus jeden Aspekt von Unternehmen betreffen. In: Der Standard, 29–30 August 2015, pp L2–L3
4. Al-Ani A, Stumpp S (2016) Rebalancing interests and power structures on crowdworking platforms. Internet Policy Rev 5(2)
5. Al-Ani A (2016) Lehren in digitalen Lernwelten. Neue Rollen und Funktionen von Lehrenden. In: Cendon E, Mörth A, Pellert A (eds) Theorie und Praxis verzahnen. Lebenslanges Lernen an Hochschulen. Waxmann, Münster, pp 247–260
6. Anderson C (2012) Makers. The new industrial revolution. Crown Business, New York
7. Bauwens M, Mendoza N, Iacomella F (2012) Synthetic overview of the collaborative economy. http://p2pfoundation.net/Synthetic_Overview_of_the_Collaborative_Economy. Accessed 10 Dec 2015
8. Benkler Y (2006) The wealth of networks. How social production transforms markets and freedom. Yale University Press, New Haven
9. Blanchet M, Rinn T, Thaden GV, De Thieulloy (2014) Industry 4.0. The industrial revolution. How Europe will succeed. Roland Berger Consultants, Munich
10. Braverman H (1974) Labour and monopoly capital. The degradation of work in the twentieth century. Monthly Review Press, New York
11. Brock C (2013) How William Shockley's robot dream helped launch Silicon valley. http://spectrum.ieee.org/at-work/innovation/how-william-shockleys-robot-dream-helped-launch-silicon-valley. Accessed 3 Dec 2015
12. Broy M (2010) Cyber-physical systems—Wissenschaftliche Herausforderungen bei der Entwicklung. In: Broy M (ed) Cyber-Physical Systems. Innovationen durch Software-Intensive eingebettete Systeme. Springer, Wiesbaden, pp 17–33
13. Brynjolfsson E, McAfee A, Spence M (2014) New world order. Labor, capital, and ideas in the power law economy. Foreign affairs, August 2014. https://www.foreignaffairs.com/articles/united-states/2014-06-04/new-world-order. Accessed 12 Dec 2015

14. Chomsky N, Rounds P, Schoeffel J (2002) Understanding power. The indispensable Chomsky. The New Press, New York
15. Deiser R (2016) Some thoughts about the future of lifelong learning universities. In: Cendon E, Mörth A, Pellert A (eds) Theorie und Praxis verzahnen. Lebenslanges Lernen an Hochschulen. Waxmann, Münster, pp 264–265
16. Florida R (2011) The rise of the creative class revisited. Basic Books, New York
17. Frazzon E, Hartmann J, Makuschewitz T, Scholz-Reiter B (2013) Towards socio-cyber-physical systems in production networks. In: Procedia of the forty sixth CIRP conference on manufacturing systems 2013, pp 49–54
18. Gorecky D (2014) Mensch-Maschine-Interaktion im Industrie 4.0-Zeitalter. In: Bauernhansl T, Hompel M, Vogel-Heise B (eds) Industrie 4.0. in Produktion, Automatisierung und Logistik. Anwendungen-technologien–Migration. Springer, Wiesbaden, pp 525–542
19. Gorz A (1989) Kritik der Ökonomischen Vernunft. Rotbuch Verlag, Berlin
20. Hagel J, Brown J, Davision L (2010) The power of pull. How small moves, smartly made, can set big things in motion. Basic Books, New York
21. Kamenetz A (2010) DIY U: Edupunks, Edupreneurs, and the coming transformation of higher education. Chelsea Green Publishing, White River Junction
22. Lohr S (2013) Computers get a human hand steering the web. New York Times. http://www.nytimes.com/2013/03/11/technology/computer-algorithms-rely-increasingly-on-human-helpers.html. Accessed 1 Apr 2015
23. Manufacturing Global (2014) Volkswagen to replace human workforce with robots in Germany to meet demand. http://www.manufacturingglobal.com/peopleskills/170/Volkswagen-to-replace-human-workforce-with-robots-in-Germany-to-meet-demand. Accessed 2 Apr 2016
24. Marglin S (1974) What do bosses do? Rev Radic Polit Econ 6(2):60–112
25. Markoff J (2015) Machines of loving grace. Harper Collins Publishers, New York
26. Moravec H (1999) Robot. Mere machines to transcendent mind. Oxford University Press, New York
27. Noble D (1985) Command performance: a perspective on military enterprise and technological change. In: Smith M (ed) Military enterprise and technological change. Perspectives on the american experience. The MIT Press, London, pp 329–346
28. Noble D (1995) Progress without people. New technology, unemployment, and the message of resistance. Between the Lines, Toronto[29] Ortmann G (1999) Formen der Produktion – Organisation und Rekursivität. Westdeutscher Verlag, Opladen.
29. Ortmann G (1999) Formen der Produktion – Organisation und Rekursivität. Westdeutscher Verlag, Opladen
30. Pfeiffer S (2015) Auswirkung von Industrie 4.0. auf die Aus- und Weiterbildung. Österreichische Akademie der Wissenschaften Verlag, Wien
31. Picot A, Neuburger R (2014) Arbeit in der digitalen Welt. Zusammenfassung der Ergebnisse der AG 1-Projetktgruppe anlässlich der IT-Gipfelprozesse 2013 und 2014. Münchner Kreis, Munich
32. Schuh G, Potente T, Varandani R, Hausberg, C, Fränken, B (2014) Variety management in manufacturing. In: Proceedings of the 47th CIRP conference on manufacturing systems collaboration moves productivity to the next level, pp 3–8
33. Shirky C (2010) Cognitive surplus. How technology makes consumers into collaborators. Penguin Books, New York
34. Siegele L (2015) Does Deutschland do digital? The Economist. http://www.economist.com/news/business/21678774-europes-biggest-economy-rightly-worried-digitisation-threat-its-industrial. Accessed 10 Dec 2015
35. The Economist (2012) Making the future. How robots and people team up to manufacture things in new ways. The Economist. http://www.economist.com/node/21552897. Accessed 2 Nov 2015
36. The White House (2015) A nation of makers. https://www.whitehouse.gov/nation-of-makers. Accessed 31 Oct 2015

37. Trudell C, Hagiwara Y, Jie M (2014) Humans replacing robots herald Toyota's vision of future. http://www.bloomberg.com/news/articles/2014-04-06/humans-replacing-robots-herald-toyota-s-vision-of-future. Accessed 2 Dec 2015
38. Vironis J (2015) It's actually open source software that's eating the world. http://venturebeat.com/2015/12/06/its-actually-open-source-software-thats-eating-the-world/?utm_content=bufferaa38e&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer. Accessed 10 Dec 2015
39. Wallerstein I (1999) The end of the world as we know it. Social science for the twenty-first century. University of Minnesota Press, London

Towards User-Driven Cyber-Physical Systems—Strategies to Support User Intervention in Provisioning of Information and Capabilities of Cyber-Physical Systems

Marko Palviainen, Jani Mäntyjärvi, Jussi Ronkainen and Markus Tuomikoski

1 Introduction

In the future there will be an increasing number of Cyber-Physical Systems (CPSs) that seamlessly integrate technologies for implementing uninterrupted interaction between the physical world and the information world for enabling human-to-human, machine-to-machine, and machine-to-human interoperation [21]. A cyber-physical system can be understood to be a network of computational systems (CPS elements) with physical inputs and outputs [16]. We believe that two kinds of CPS will exist: (a) fully automatic CPSs and (b) user-driven CPSs. Fully automatic CPSs are systems that automatically exploit available CPS elements and serve their users in their everyday life. An example of a fully automatic CPS is a cyber-physical system that optimizes energy efficiency at home and tries to minimize peak demand for electricity and thus reduces greenhouse gas emissions. We believe that cyber-physical systems will not consist of a fixed set of CPS elements but will rather evolve over time; the user will have a full freedom to add different kinds of elements to cyber-physical systems, to remove elements from them, and to affect the provisioning of information and capabilities of cyber-physical systems. The diversity of cyber-physical systems raises

M. Palviainen (✉)
VTT Technical Research Centre of Finland Ltd, P.O. Box 1000, 02044 Espoo, Finland
e-mail: marko.palviainen@vtt.fi

J. Mäntyjärvi · J. Ronkainen · M. Tuomikoski
VTT Technical Research Centre of Finland Ltd, P.O. Box 1100, 90570 Oulu, Finland
e-mail: jani.mantyjarvi@vtt.fi

J. Ronkainen
e-mail: jussi.ronkainen@vtt.fi

M. Tuomikoski
e-mail: markus.tuomikoski@vtt.fi

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_24

a new question: How can the user exploit the full potential of the available CPS elements in a certain use context or in a certain task? One possibility to solve the problem is a *user-driven CPS* that partially moves the burden of cyber-physical system, application and service development from software professionals to the end users of the systems. A *user-driven CPS* enables the end user to affect the exploitation of the information and capabilities of cyber-physical systems, both within and outside of the CPSs, in applications and services that are based on the information and capabilities of the CPSs. For example, a cyber physical production system in a factory could allow the use of the monitoring and controlling capabilities of the factory machines for end user development. Thus, a new CPS display element that was installed to the wall of the factory could be utilized by a service person who would create a small application that fetches monitoring information from the machines and service instructions from a maintenance service, and show them on the new CPS display.

The user-driven CPS approach requires user intervention solutions that enable the users to: (a) affect the behaviour of a CPS and the applications and services built around it, and (b) develop new applications and services that are based on information and capabilities of CPSs. The user intervention solutions depend (at least) on: (a) *users' motivation* i.e. some users just want to use readily available applications and services, some are interested in reconfiguring applications and services for different purposes, and some want to develop new applications or services for their needs, (b) *users' skills,* i.e., the end user can be a software professional, a software enthusiast or a non-programmer and (c) *underlying CPS architecture and implementation* i.e., the architectural choices and the capabilities of the used CPS elements will affect the development of services.

When designing solutions that enable users to intervene with a cyber physical system one should keep in mind the general principles for design and usability [15]: *Visibility*, interactors should be clearly visible and they should, by design, explain their purpose as well as possible. *Feedback*, providing information immediately back to a user after (s)he has done something with an interactor. *Constraints*, in order to avoid confusion or misunderstanding, the user's focus should only be on the desired interactor. Thus, one should be able to limit the interaction possibilities for a given interaction. *Mapping*, describes the relationships between controls and their effects with the real world. With a complex cyber-physical system this is especially challenging since the CPS device that is controlled may not be physically present but can be located elsewhere. *Consistency*, the interaction elements and paths between different tasks when used by a same "device" or interaction environment must be similar in style, logic and operation (e.g., response and feedback). This promotes learning and acceptance.

The goal of this chapter is to outline strategies for enabling user intervention: (a) in the behaviour of CPS and (b) in the behaviour of services that support use of cyber-physical systems. In addition, an architecture is outlined for user-driven CPSs that offer support for different kinds of user groups, and for the proposed user intervention strategies.

The remainder of this chapter is organized as follows. The first section introduces background related to cyber-physical systems, CPS architectures, and solutions that enable user intervention in the provisioning of the information and capabilities of the cyber-physical system. The second section outlines a Cyber Physical Production System to exemplify the concept of a user-driven CPS. The third section introduces strategies for user intervention in the behaviour of a CPS and related services. The fourth section outlines an architecture for user-driven cyber-physical systems that exploit the proposed user intervention strategies. The benefits and challenges related to the presented strategies are discussed in the fifth section. Conclusions are finally drawn in the sixth section.

2 Background

This section introduces background related to cyber-physical systems and solutions supporting creation of user-driven CPSs.

2.1 Cyber-Physical Systems

A cyber-physical system features a tight combination of coordinated computational and physical elements [8]. A cyber-physical system integrates computation with physical processes that affect computations and vice versa [9]. The embedded systems in aerospace, automotive, chemical processes, civil infrastructure, energy, healthcare, manufacturing, transportation, entertainment, and consumer appliance areas can be seen to be a precursor generation of CPSs [8]. However, unlike traditional embedded systems, a full-fledged CPS is typically designed as a network of interactive elements (see Figs. 1 and 2) rather than standalone devices [8], i.e., a CPS represents not only a single device, but also a collection of devices working in concert [16].

Ongoing advances in science and engineering are assumed to improve the link between computational and physical elements, significantly increasing the adaptability, autonomy, efficiency, functionality, reliability, safety, and usability of cyber-physical systems [8]. The physical elements of CPSs may include sensors and actuators such as photocells, physical buttons, motors, and displays or LEDs [16].

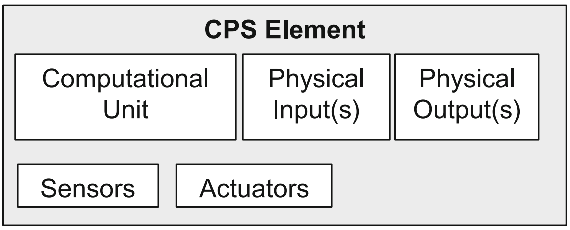


Fig. 1 A model for a CPS element

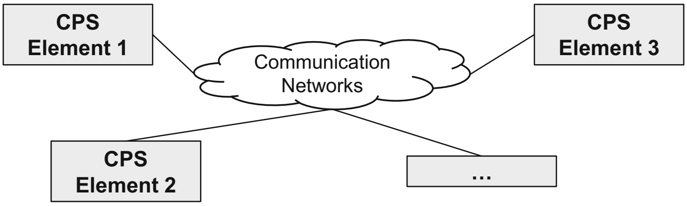


Fig. 2 A model for a cyber-physical system

From a computational viewpoint, CPSs can also be viewed as computational systems with inputs and outputs in the physical world [16]. Viewing the CPS elements as computational elements allows the potential application of software engineering concepts, such as object-oriented development and high-level process architectures to physical machines, in addition to the logical machines of software [16].

2.2 Solutions for User Intervention in Cyber-Physical Systems

The following paragraphs discuss approaches that could be used for user intervention in the behaviour cyber-physical systems and in the behaviour of applications and services that are based on cyber-physical systems.

End user Development (*EUD*)—The end user knows best what kinds of application and services (s)he needs. End user Development (EUD) consists of methods, techniques, and tools that allow users of software systems to act as non-professional software developers that create, modify, or extend software artefacts [11].

End user Programming (*EUP*) is programming that any computer user can perform. In contrast to professional programming, the target of EUP is to produce a program primarily for personal rather than public use, i.e., programs produced in EUP are not primarily intended for a large number of users [7].

End user development of mashups—EUP and mashups are identified as enablers for managing data, presentation and functionalities in the Internet-of-Things (IoT) context [20]. Internet of Things will create huge amounts of distributed streaming data which can be modelled using the RDF data model called *Linked Stream Data* and used as a source for *IoT mashups* [19]. In the vision of *enterprise mashups* [6] the mashup approaches are ported to company intranets, enabling enterprise members to play with a company's internal services that give access to enterprise information assets, and to mash them up in innovative, value-generating ways, for example, to automate recurrent bureaucratic procedures. In the enterprise context it is indeed possible to recognize two main situations [2]: (a) expert developers use mashup tools for delivering applications quickly for end users that are not directly involved in the construction of mashups but benefit from the shorter turn-around time for new applications. (b) Expert developers create services in a format that can be more easily consumed and combined into mashups by users who are not themselves developers, for example requiring simple parameterizations of components; they also provide a tool where anyone creates their own mashups.

Tools for user-driven EUD—There are solutions supporting the creation of reactive applications for information sources such as Event-Condition-Action (ECA) rule matching service [12, 1] that enable the chaining of simple conditional statements. EUP tools for ubiquitous environments are focused on streamlining the use of input languages or metaphor-based graphical user interfaces, aimed at simplifying the use of applications for the users [2]. Event-based and rule-based approaches such as Pervasive interactive Programming (PiP). Chin et al. [3] have also been developed for enabling end users to customize their personal space in ubiquitous environments.

2.3 Architectures for Cyber-Physical Systems

The following paragraphs introduce different kinds of architectural models for cyber-physical systems.

Service-oriented architectures for CPSs—The service-oriented architecture framework is a three layer architecture for CPSs that consists of a physical layer, a network layer and a service layer [21]. WebMed is a service-oriented middleware architecture for cyber-physical systems that supports the utilization of physical devices, computing elements and other software services. It pays particular attention to bringing the underlying physical devices' capabilities directly to the application development layer, and linking them with non-physical device services using service orientation principles such as loose coupling, repository and discovery, reuse, and composition of services [4]. The dynamic virtual environment architecture is also developed to support service-oriented control methods and to enable the users to control appliances in the physical environment by intuitive operations through a virtual (home) environment on the network [8].

Event-based architecture for CPSs—An event-action-based and hierarchical architecture for CPS separates events to separate layers and supports the representation of composite event conditions [17].

REpresentational State Transfer (REST)-style architecture for CPSs—The REST-style architecture for CPSs contains a smart gateway that is based on the REST architectural style and offers a uniform, efficient, and standardized way of interacting with physical devices [10]. As a benefit, applications can be built with little effort on top of heterogeneous devices with any programming language that supports HTTP. However, the REST-style architecture is based on the pull-based interaction model that is not suitable for modelling the communication with smart devices with respect to monitoring.

Architecture for cross-smart space applications—We have previously developed an architecture for end user programming of cross-smart space applications (described in detail in [13, 14]) that enables end user development of applications that are based on capabilities and information of different smart spaces. In the architecture a smart space is understood to be a Semantic Information Broker (SIB), a lightweight (RDF) database that: (a) takes care of information storing, sharing and governing and (b) provides add, remove, query, and subscribe functions for the users of SIB and for the semantic information stored to the SIB [5, 18]. The cross-smart space application architecture is based on multiple blackboards, i.e., smart spaces that enable information exchange between entities that are connected to the smart spaces. A similar architecture could be used in user-driven cyber-physical systems, too. The CPS elements could connect to smart spaces to share their information and capabilities with tools that enable user intervention in the behaviour of a cyber-physical system, and in the behaviour of applications and services that are based on the cyber-physical system.

2.4 Requirements of User Intervention for Cyber-Physical System Architecture

Implementation of user intervention for CPS sets special requirements for architecture. First of all, **underlying architectural and implementation issues should be hidden from the user**. It should be enough that the user focuses on user intervention without having to think about reflections of the architectural or implementation issues of the underlying cyber-physical system. Secondly, **user intervention with the system should be (or forced to be) usable in a controlled manner.** User intervention changes the state of CPS. The user is typically accustomed to interacting with other simple applications and user interfaces. User intervention with a CPS should preferably be similar to interaction with other systems. Moreover, the intervention should occur in a controlled way in run-time without causing harms for the CPS, its users, or the physical phenomena it monitors or controls. For example, user intervention should not cause lost of important information in a CPS, and the user intervention should not put the CPS to an erroneous or harmful state. Moreover, new user interfaces should be created in a consistent manner to ensure consistency and clarity. Thirdly, **security mechanisms for user intervention**—for security reasons it must be ensured that only authorized users perform user interventions in CPS and supporting services and applications. It should also be ensured that interventions the users perform are authorised.

3 Example—A User-Driven Cyber Physical Production System

A user-driven CPS can consist of CPS elements and supporting services. CPS elements can provide controlling, monitoring, and configuration capabilities. A *supporting service* can be an application or service that supports the use of a

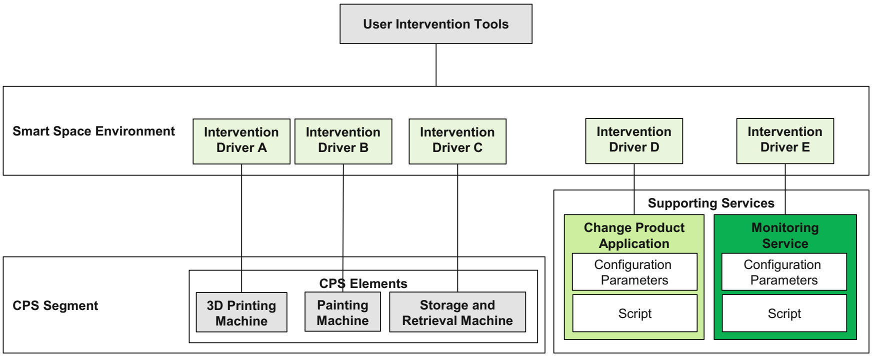


Fig. 3 An example of a user-driven CPPS

cyber-physical system and can, for example, provide capabilities for information fetching, processing, exploitation, and visualisation.

Figure 3 depicts an example of a cyber physical production system that is a user-driven CPS and enables user intervention in the behaviour of the elements of the system. The following paragraphs describe the most important elements of the user-driven CPPS example.

CPS elements—The outlined CPPS consists of three CPS elements:

(1) **3D Printing Machine**—A machine that can be configured to produce 3D items from different materials.
(2) **Painting Machine**—The painting machine paints produced items with desired colours.
(3) **Storage and Retrieval Machine**—The storage machine is capable of packaging produced items, storing produced items to the storage and retrieving items from the storage.

Supporting services—The example CPPS provides an application and service for end users. Firstly, the end-user can use the Change Product Application for configuring the CPPS to produce desired items with the defined quality-level. Secondly, the Monitoring Service is a Web service that fetches information from different CPS elements and provides monitoring information about the amount of different materials available for 3D printing and about the status of the storage of the produced items.

CPS Segment—User intervention is based on a *CPS segment* that defines the target of the user intervention. In the example three elements of the cyber physical production system are included in the CPS segment.

Intervention driver—Our end user programming approach for cross-smart space applications [13, 14] uses software components (drivers) for providing capabilities of software systems to be exploited in end user programming. A similar kind of approach can be used in a user-driven CPS by introducing *intervention drivers* that enable the use of the capabilities and information available via the APIs of CPS elements and supporting services.

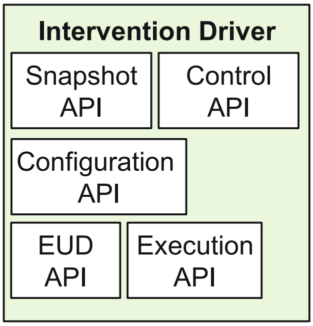


Fig. 4 An intervention driver

An intervention driver in Fig. 4 can provide following kinds of modules for user intervention:

1. **Snapshot API—**This module enables fetching of information for user intervention. For example, this information can contain user interface descriptions for UIs that enable interaction with the CPS elements, configuration parameters that assist the end user to set the CPS elements to the default states or to desired modes, and readily made scripts that can be available in the CPS elements.
2. **Control API**—This module enables direct controlling of CPS elements via available user interfaces.
3. **Configuration API**—This module enables remote configuration of the CPS element. For example, the 3D printing machine can be configured to produce desired items from the defined materials.
4. **Execution API**—This module enables remote execution of commands in the CPS elements.
5. **EUD API**—This module enables the use of the capabilities and information of CPS elements and their supporting services for end user development. An EUD API can be a Semantic End-User Application Programming Interface (S-API) [13, 14] that defines the driver's (a) commands (i.e. driver's capabilities), (b) inputs (i.e. information provided by the driver), (c) outputs (i.e. information consumed by the driver), and (d) execution branches that relate to the commands. For example, the EUD API of the 3D printing machine could provide a "Get Surface Quality" command that provides an output to be used as an input for the painting machine's "Set Painting Parameters" command.

User intervention tools—The user intervention tools are capable of: (a) handling CPS segments and (b) performing centralized user intervention that affects the behaviour of CPS segments and exploitation of the capabilities and information of the elements of CPS segments.

Smart space environment—The user intervention tools and intervention drivers are joined to the smart spaces and can publish information to the smart spaces, query information from the smart spaces and subscribe information from the smart spaces, i.e., monitor what kind of information is published to the smart spaces. The use of the APIs of intervention drivers can be based on the publish-subscribe mechanisms of the smart spaces that enable the user intervention tools to send requests to the drivers and receive responses from them.

4 Strategies for User-Driven Cyber-Physical Systems

This section outlines a process and strategies that enable user intervention in the behaviour of cyber-physical systems and supporting services. The user intervention process consists of four steps (depicted in Fig. 5):

(1) **Selection of target elements**—This step produces a *CPS segment* that defines the target of the user intervention. Cyber-physical system elements and supporting services can be selected to be included in CPS segments.
(2) **Preparation for user intervention**—The preparation of user intervention can require deactivation of the elements of the CPS segment, or it may suffice to take a snapshot of the elements of the CPS segment that is used as a base for the user intervention. The information needed in the snapshot depends on the type of user intervention. For example, UI descriptions, configuration scripts, and EUD APIs can be included in snapshots.
(3) **User intervention**—User intervention encompasses end user actions that at some level aim at affecting the behaviour of the elements of the CPS segment. The step produces a user intervention description for the needed behaviour changes.
(4) **Deployment of user intervention**—This step takes the user intervention description as an input, possibly validates the correctness of the described user intervention action, possibly deactivates the target elements of the user intervention, deploys the accepted changes to the target elements, and finally activates the target elements of the user intervention.

The rest of this section focuses on the user intervention step. The next subsections outline strategies that enable user intervention in provisioning of information and capabilities of cyber-physical systems and supporting services.

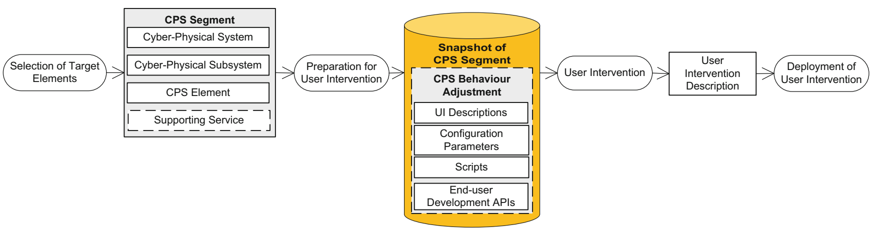


Fig. 5 A process for user intervention in CPS and in supporting services

4.1 Strategies for User Intervention in Behaviour of CPS

The following paragraphs introduce different kinds of strategies to affect the provisioning of information and capabilities within a CPS.

Firstly, the end user may be interested in modifying the behaviour of the entire cyber-physical system, its subsystems, or particular CPS elements only. This requires strategies for: (a) *centralized user intervention* and (b) *element-centric user intervention.*

Centralized user intervention—Centralized user intervention means a coordinated exploitation of the elements in a CPS or its subsystems. An example of a centralized user intervention is a user-initiated reset of CPS that sets CPS elements to their initial state. However, the reset of CPS can be difficult to implement in practice because CPS reset must be performed in a controlled way, and without losing any important information. In addition, network disconnections and malfunctions of CPS elements can prevent the remote reset of CPS elements.

Centralized user intervention requires CPS element search and selection mechanisms to produce CPS segments. The production of CPS segments can be based on: (a) end user's interests, (b) state of CPS elements, i.e., the search can identify elements that need user intervention, (c) user profile and user preferences, (d) user task (e.g., searches elements that are needed in the end user's task) and (e) usage context (e.g., searches elements that are needed in the identified use context). Centralized user intervention also requires centralized mechanisms for affecting the behaviour of the elements of CPS segments and for the exploitation of their capabilities and information.

Element-centric user intervention—An element specific user intervention means a case in which the end user focuses on affecting the behaviour of one CPS element at a time. An example of an element-centric user intervention is a situation in which the end user uses a configuration tool to modify the configuration parameters of the selected element. Another example could be the use of rule-based CPS elements combined with a rule editor that enables the end user to define how the selected element reacts to events in the surrounding environment. Secondly, there are (at least) three ways to perform user intervention in CPS segments: (a) direct interaction with CPS segments, (b) parametric configuration of CPS segments, and (c) execution of scripts.

Direct interaction with CPS segments—In direct interaction with CPS segments the end user uses available monitoring methods and controls to immediately see the changes in the behaviour of the CPS segment. The snapshot of a CPS segment can contain *UI descriptions* for UIs that enable direct interaction with the elements of the segment. For example, a UI description can define a user interface that a) provide information about the state changes in the elements of CPS segments for the end user and b) enable the end user to activate or deactivate CPS or its parts, or reset the CPS to an initial or otherwise safe state.

Parametric configuration of CPS segments—Often the user intervention in a CPS is just fine-tuning; it is enough that small details of the CPS are configured for the needs of the end user. This kind of fine-tuning can be based on a set of configuration parameters and a mechanism that updates the configuration parameters to the elements of CPS segments. The elements can provide predefined configuration parameters that assist the end user to set the elements to the default states or to desired modes. These configuration parameters can be delivered in the snapshot of a CPS segment.

Execution of scripts—In this solution, scripts are used to define a set of operations that can fetch information from CPS elements, produce configuration parameters based on the fetched information, and finally set the system to the desired state i.e. modify the behaviour so that CPS will better serve the needs of the end user. The snapshot of a CPS segment can contain readily made scripts that are available in the elements of the segment. For example, there could be a script for the CPPS outlined in Sect. 3 that fetches information about surface-quality from the 3D printing machine and then sets the painting parameters of the painting machine based on that information. Execution of scripts can be based on centralized or distributed execution, in which case execution is performed remotely in the elements (in CPS elements and supporting service modules) of the CPS segment.

4.2 Strategies for User Intervention in Behaviour of Supporting Services

The following paragraphs introduce solutions that enable the end user to affect the provisioning of information and capabilities in applications and services that are based on CPS.

Parametric configuration of supporting services—This is needed in a case in which it is enough that small details of supporting services are configured for the needs of the end user. This kind of fine tuning can be based on reconfiguration of parameters of a service to set it to a desired state.

End user development of supporting services—In the computer world the default approach is to write applications or services that enable the end user to exploit the available capabilities and information. In the case of CPS, there is a great need for end user development because it is very difficult to write ready-made applications or services to exploit the capabilities and information of all possible element configurations.

End users who are motivated to create applications and services for their everyday life need end user programming tools that decrease the complexity of application and service development and abstract the architectural and implementation details of the underlying cyber-physical system. In practice, certain kinds of elements must be assumed to exist to provide the required capabilities and information for application or service development. In EUD the end user uses appropriate tools and selects the CPS elements and supporting services to be used in EUD. The selection of CPS elements and supporting services can be automatized and, for example, search mechanisms may be available to assist searching of

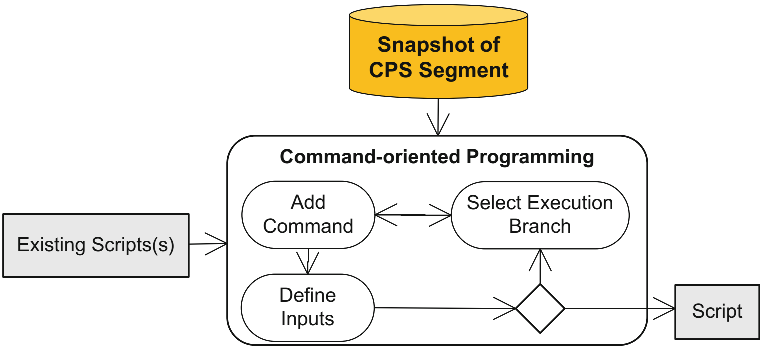


Fig. 6 Command-oriented programming [14] of scripts

capabilities or information for the end user's tasks. This step produces a CPS segment that provides a starting point for end user development.

Subsequently, a snapshot is taken that will now contain the EUD APIs of the intervention drivers of the elements of the CPS segment. The end user uses an appropriate EUD tool that reads the EUD APIs from the snapshot and enables the end user to create a new script (an execution sequence) that defines the operations performed in the created application or service, and the information that is produced and exploited in the script. Creation of scripts can be based on existing scripts (if development is not started from scratch) and the command-oriented-programming [14] method that consists of three main steps: (a) Select Execution Branch, (b) Add Command, and (c) Define Inputs (see Fig. 6). The available commands can be obtained from the EUD APIs in the snapshot of the CPS segment.

The command-oriented programming produces a script that can be executed in the target system. The script defines a command sequence and information chains between the outputs and inputs of the commands. This means that the produced script will define the operations that are executed in the script and the information that is produced and exploited in the script.

Extension of supporting services—In this case, the end user has an application or service that (s)he desires to extend for his/her needs. For example, the end user may have a new CPS element (e.g., a smartwatch) that provides user interface (e.g., display and control) capabilities that could improve the user experience of an existing service. The end user performs now end user development to extend the service to use the capabilities and information of the new CPS element.

Reduction of supporting services—The end user has a supporting service that offers important features but also features that he or she does not actively use. So, for improving usability, the end user hides these unnecessary features from the supporting service using EUD tools.

5 Architecture for a User-Driven Cyber-Physical System

This section outlines an architecture for user-driven CPS (depicted in Fig. 7) that aims at supporting the user intervention strategies introduced in Sect. 4. In the architecture, two kinds of users for the user-driven CPS are described: end users and CPS developers. An end user can be a non-programmer, software enthusiast, or software professional that uses the available user intervention mechanisms. In the case of the CPPS example in Sect. 3 the end users could be maintenance personnel who ensure the continuous operation of the production process in the CPPS. A CPS developer is a software professional or software enthusiast that implements intervention drivers for user intervention.

In order to provide the functionalities required in the user intervention strategies described in Sect. 4, the architecture must outline modules and tools to enable user intervention in the behaviour of the CPS elements and supporting services. The architecture for a user-driven CPS consists of a smart space environment, CPS developer tools, and user intervention tools. The CPS developer tools assist the development of intervention drivers for the APIs of CPS elements and for the APIs of supporting services. The user intervention tools enable the end users to affect the behaviour of the elements of cyber-physical system and supporting services.

The following subsections discuss the architecture form two viewpoints, from: (a) the End user viewpoint and (b) the CPS developer viewpoint.

5.1 End User Viewpoint

The following paragraphs outline modules that are needed to enable user intervention in the behaviour of cyber-physical system, cyber-physical subsystems, and CPS elements:

(1) *Modules for centralized user intervention and CPS element centric user intervention*—The CPS Element Management Module is capable of searching and selecting of CPS elements, producing CPS segments for user intervention and taking snapshots of CPS segments. The CPS Element Management

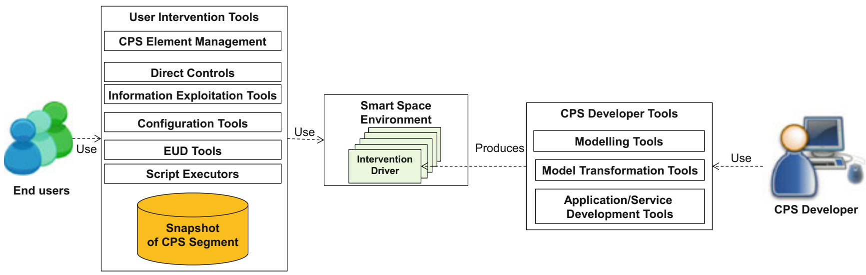


Fig. 7 The architecture for a user-driven cyber-physical system

module enables fine-grained selection of CPS elements and enables the end user to manually select the CPS elements for user intervention.

(2) *Modules for direct interaction with cyber-physical systems*—The Control APIs, Snapshot APIs and the user interfaces of the intervention drivers enable the end user to monitor the state of the elements in CPS segments and to control them. Examples of direct controls related to CPS segments are element activation, deactivation, and reset.

(3) *Modules for monitoring of cyber-physical systems*—Snapshot APIs and information exploitation tools do not try to control the CPS segments but rather monitor the state of the elements of CPS segments or support exploitation of their information. For example, like in the Internet world, there could be browsers or portals for information available in CPSs or then there could be visualization environments that enable the end users to exploit the information available in CPS segments.

(4) *Modules for parametric configuration of cyber-physical systems*—Often it is enough that the end user is capable of performing fine-tuning for CPS segments. The intervention driver can provide configuration APIs for configuration tools that enable the end user to change the configuration parameters of the elements of CPS segments in a controlled way.

(5) *Modules for end user development of scripts*—The CPS Element Management module enables the selection of CPS elements for EUD and produces a respective CPS segment. The EUD tool reads the EUD APIs from the CPS segment snapshot and enables end user development of application or service scripts.

(6) *Modules for execution of scripts*—Execution of scripts can be based on a central entity (Script Executor) that orchestrates the execution of scripts in CPS segments. However, it is also possible that the creation of a supporting service is based on centralized development but the end result is executed in distributed remote script executors that are available in the intervention drivers. An example of this kind of system is a cyber-physical system that: (a) enables the end user to define rules for CPS element behaviour in centralized way and then (b) transfers the defined rules to the CPS elements that then take care of execution of the rules.

The following paragraphs outline APIs and tools that are needed to enable user intervention in the behaviour of supporting services:

(1) *Modules for parametric service configuration*—The configuration APIs of intervention drivers are needed to enable the changing of configuration parameters of supporting services.

(2) *Modules for end user development*—End user development can be based on command-oriented programming [14] and an EUD tool that enables the end user to develop a script that defines how an application or service is composed in the target system.

(3) *Modules to enable execution of end user's applications or services*—The execution APIs and script executors are needed to enable execution of end user's application or service scripts.

5.2 CPS Developer Viewpoint

The development of intervention drivers requires domain knowledge. For example, the CPS developer should know the needs of the end users of a particular domain and identify the intervention drivers and define the APIs (e.g. Snapshot API, Control API, Configuration API, Execution API, and EUD API) that should be implemented to support user intervention in the CPS segments.

Our previous experience of driver development [13, 14] shows that the development of intervention drivers should be based on co-creation and iterative development and test cycles. The CPS developer should first create initial intervention drivers for user intervention. End users will then test these modules and provide feedback for the CPS developer. Subsequently, after multiple development and test cycles, intervention drivers should be good enough for widespread use. For example, opening of CPS element's capabilities and information for EUD requires that provided EUD APIs (e.g., naming of operations and their inputs and outputs in APIs) are good enough for the end users. In addition, the development of data models for the input and output information of CPS elements requires domain knowledge about information that is produced and/or exploited in application or service scripts.

Efficient development methods and tools are needed to minimise the implementation effort of intervention drivers. It can be outlined that the development of intervention drivers for user intervention requires tools for:

(1) **Modelling**—Modelling tools are needed for:

  a. **Controls modelling**—There should be tools for modelling and describing the direct control mechanisms for CPS elements.
  b. **Configuration parameter modelling**—There should be tools for modelling the configuration parameters of CPS elements.
  c. **Capability modelling**—End user development requires information about the capabilities that can be used in end user's application/service scripts. Thus, there should be tools for modelling the capabilities of CPS elements.
  d. **Information modelling**—End user development requires data models for information that is produced or exploited in end user's application/service scripts. Thus, there should be tools for modelling information that CPS elements provide as an output and/or exploit as an input.

(2) **Model transformations**—There should be tools capable of transforming code (e.g., code skeletons for intervention drivers) from the models defined in the modelling tools to assist implementation of intervention drivers that can

implement Snapshot APIs, Control APIs, Configuration APIs, Execution APIs, and EUD APIs.

(3) **Reference application creation**—Reference applications and services are needed to guide the end users in the exploitation of intervention drivers. The developers can use the end user development tools in creation of reference applications/services for the drivers.

In addition to tools, a distribution channel is required for intervention drivers, too. There should be an ecosystem that enables developers to: (a) distribute intervention drivers and supporting services for different kinds of end users, (b) collect feedback from end users and (c) share elements (e.g., models) used in the development of intervention drivers for other actors in the ecosystem. However, the CPS ecosystem and distribution of user intervention elements issues are not in the scope of this chapter and thus are not discussed in more detail here.

6 Discussion

The exploitation of the strategies outlined in Sect. 4 depends on the end user's motivation, skills and needs. Sometimes it is enough that just fine-tuning is performed for CPS and/or supporting services but sometimes there is a need for end user development or even coding. In addition, the end users for cyber-physical systems vary. For example, there are: (a) end users who are just interested in using cyber-physical systems, (b) end users who want to configure CPS elements or supporting services for different purposes and (c) end user that are interested in to create new services, and (d) end users such as software professionals and software enthusiasts who may want to do programming for extending a cyber-physical system for new purposes and needs. The proposed architecture outlines modules that are needed to provide support for the CPS developers and for the end users of cyber-physical systems.

The following lists the most important benefits related to a user-driven CPS:

Can make end users' everyday life easier—The user-driven CPS enables end user development of applications and services in the intersection of CPS elements and in the digital services in the Web. In addition, it is possible to create execution flows and information flows between CPS elements and digital services. This can benefit end users, i.e., they could exploit the CPS elements (e.g., as controls and displays of digital services) in the use of digital services, and vice versa.

Provides additional value for CPS elements and digital services—A user-driven CPS can enrich the exploitation of the capabilities and information of the elements (e.g., physical devices) of cyber-physical systems and digital services that can be combined to interoperate with CPS elements. An example of digital service that could be exploited in controlling CPS elements is a service that provides electricity price information in the Web. This information could be combined to control the use of electricity in CPS elements, for example to control the CPS elements controlling the heating and air conditioning in a building. An example of an opposite mechanism could be a thermostat in a CPS that can predict when the heating will start in the building and act as an information source for a digital service. In end user development, the end user can define an application that delivers this information for a digital service of an electricity producer that could now better optimise the electricity production power for the forthcoming demand of electricity.

The following lists the most important challenges related to user-driven CPSs:

Need for CPS elements' APIs—The implementation of an intervention driver is possible only if a CPS element provides the necessary APIs for using the CPS elements' capabilities and information in the intervention drivers.

Need for intervention drivers—Developing intervention drivers for the APIs of CPS elements and digital services require software professionals or software enthusiasts. In practice, an ecosystem and business models need to be available for the modules of user-driven CPSs.

Performance—The fluent use of a user-driven CPS requires good performance from the used network connections, CPS elements, supporting services, and execution modules that manage the execution of end users' application or service scripts.

Reliability—Handling the complexity of a cyber-physical system is an issue and it can be difficult to ensure the reliability of a user-driven CPS because there are many issues affecting the reliability of the system. For example, unreliable network connections, intervention drivers, or execution modules and erroneous applications and services produced in end user development can negatively affect the reliability of user-driven cyber-physical systems.

The following discusses how general requirements of the CPS architecture introduced in Li et al. [10] are considered in the outlined user-driven CPS architecture:

- **Composition**—The heterogeneity of cyber-physical systems will lead to systems that don't behave well outside of a small operational envelope and that are hard to maintain. The outlined architecture introduces modules for end user development that enables the end users to create applications and services for heterogeneous and changing cyber-physical systems. However, end user development is required if the CPS elements that are used in the end user's applications/services are removed from the cyber-physical system.
- **Systems integration**—Since CPSs couple physical processes with computing and communication processes, system component design becomes complex. With current technology, it is difficult to integrate complex components into complex systems. The proposed architecture supports end user driven integration of CPS elements' and supporting services' capabilities and information. However, it is required that intervention drivers are provided for the used CPS elements and supporting services.
- **Reliability**—Cyber-physical systems will not be operating in a controlled environment, and must be robust against unexpected conditions and adaptable to

subsystem failures [9]. Since physical devices interface with the uncertainty and the noise in the physical environment, systems must tolerate or contain component failures in both the cyber and the physical domains. End user development can be based on the command-oriented-programming method that enables the end users to define recovery mechanisms for exceptional situations. The used capabilities can define error branches to which the end user can connect error handling commands to handle exceptional situations. However, there is still a need for further work enhancing the failure handling mechanisms of the proposed architecture.
- **Security and Privacy**—Cyber-physical systems open up new threats: physical systems can now be attacked through cyberspace and cyberspace can be attacked through physical devices. The proposed architecture is based on the smart space architecture model that assumes that only trusted and authenticated entities may join to the smart spaces. However, there is still a need for more advanced methods for managing security and privacy issues in smart spaces. One more issue is that since user intervention changes the state of CPS elements, it should be performed in a controlled way during run-time without causing harm to the CPS or its users. For example, it must be ensured that important information is not lost in the CPS during user intervention, and that user intervention cannot put the CPS into an erroneous or harmful state.

7 Conclusions

This chapter outlines strategies for enabling user intervention in the behaviour of cyber-physical systems and supporting services. In addition, an architecture that offers support for the exploitation of the proposed user intervention strategies is outlined for user-driven cyber-physical systems.

The development of a user-driven cyber-physical system requires co-creation and rich communication between the end users and the developers of cyber-physical systems. Especially, the naming of capabilities and information elements strongly affects the usability of user intervention tools. Thus, it is very important that both the capability models and information models used in intervention drivers are iteratively developed in tight co-operation with end users.

References

1. IFTTT. https://ifttt.com. Accessed 5 Apr 2016
2. Cappiello C, Daniel F, Matera M, Picozzi M, Weiss M (2011) Enabling end user development through mashups: requirements, abstractions and innovation toolkits. In: Costabile MF, Dittrich Y, Fischer G, Piccinno A (eds) End-user development. Lecture notes in computer science, vol 6654. Springer, Berlin, pp 9–24

3. Chin J, Callaghan V, Clarke G (2006) An end user tool for customising personal spaces in ubiquitous computing environments. In: Zhang D, Portmann M, Tan AH, Indulska J (eds) Ubiquitous intelligence and computing. Lecture notes in computer science, vol 4159. Springer, Berlin, pp 1080–1089
4. Hoang DD, Paik H-Y, Kim C-K (2012) Service-oriented middleware architectures for cyber-physical systems. Inter J Comput Sci Netw Secur 12(1):79–87
5. Honkola J, Laine H, Brown R, Tyrkko O (2010) Smart-M3 information sharing platform. In: IEEE symposium on computers and communications (ISCC), Riccione, Italy, 22–25 June 2010, pp 1041–1046
6. Jhingran A (2006) Enterprise information mashups: integrating information, simply. In: Proceedings of the 32nd international conference on very large databases (VLDB), Seoul, Korea
7. Ko AJ, Abraham R, Beckwith L, Blackwell A, Burnett M, Erwig M, Scaffidi C, Lawrance J, Lieberman H, Myers B, Rosson MB, Rothermel G, Shaw M, Wiedenbeck S (2011) The state of the art in end-user software engineering. ACM Comput Surv 43(3):1–44. doi:10.1145/1922649.1922658
8. Lai C-F, Ma Y-W, Chang S-Y, Chao H-C, Huang Y-M (2011) OSGi-based services architecture for cyber-physical home control systems. Comput Commun 34(2):184–191
9. Lee EA (2008) Cyber physical systems: design challenges. In: 11th IEEE international symposium on object oriented real-time distributed computing (ISORC), 5–7 May 2008, pp 363–369
10. Li Q, Qin W, Han B, Wang R, Sun L (2011) A case study on REST-style architecture for cyber-physical systems: restful smart gateway. Comput Sci Inf Syst 8(4):1317–1329
11. Lieberman H, Paterno F, Klann M, Wulf V (2006) End-user development: an emerging paradigm. In: Lieberman H, Paternò F, Wulf V (eds) End user development, vol 9. Springer, Netherlands, pp 1–8
12. López-de-Ipiña D (2001) An ECA rule-matching service for simpler development of reactive applications. In: Middleware 2001, Heidelberg, Germany, November 2001
13. Palviainen M, Kuusijärvi J, Ovaska E (2012) Framework for end-user programming of cross-smart space applications. Sensors 12(11):14442–14466
14. Palviainen M, Kuusijärvi J, Ovaska E (2014) A semi-automatic end-user programming approach for smart space application development. Pervasive Mob Comput 12:17–36
15. Preece J, Sharp H, Rogers Y (2002) Interaction design: beyond human-computer interaction. Wiley, Chichester
16. Robbins DE, Tanik MM (2014) Cyber-physical ecosystems: app-centric software ecosystems in cyber-physical environments. In: Suh SC, Tanik UJ, Carbone JN, Eroglu A (eds) Applied cyber-physical systems. Springer, New York, pp 141–147
17. Tan Y, Vuran MC, Goddard S (2009) Spatio-temporal event model for cyber-physical systems. In: 29th IEEE international conference on distributed computing systems (ICDCS) workshops, 22–26 June 2009, pp 44–50
18. Toninelli A, Pantsar-Syväniemi S, Bellavista P, Ovaska E (2009) Supporting context awareness in smart environments: a scalable approach to information interoperability. In: The international workshop on middleware for pervasive mobile and embedded computing, Urbana Champaign, Illinois. ACM. doi:10.1145/1657127.1657134
19. Tran TN, Truong DK, Hoang HH, Le TM (2014) Linked data mashups: a review on technologies, applications and challenges. In: Nguyen NT, Attachoo B, Trawiński B, Somboonviwat K (eds) Intelligent information and database systems. Lecture notes in computer science, vol 8398. Springer International Publishing, Heidelberg, pp 253–262
20. Uckelmann D, Harrison M, Michahelles F (2011) An architectural approach towards the future internet of things. In: Uckelmann D, Harrison M, Michahelles F (eds) Architecting the internet of things. Springer, Berlin, pp 1–24
21. Xiao-Le W, Hong-Bin H, Su D, Li-Na C (2011) A service-oriented architecture framework for cyber-physical systems. In: Qian Z, Cao L, Su W, Wang T, Yang H (eds) Recent advances in computer science and information engineering. Lecture notes in electrical engineering, vol 126. Springer, Berlin, pp 671–676

Competence Management in the Age of Cyber Physical Systems

Peter Letmathe and Matthias Schinner

1 Companies in the Age of Industrie 4.0

Production in high-wage countries is being increasingly influenced by an aging workforce and the explosion of knowledge [1, 21]. As this aging workforce retires, knowledge will also leave the companies [16]. Furthermore, a higher number of more complex products, new production processes, growing competition through internationalization, and especially new technologies in markets with rapidly changing conditions are a tremendous challenge for companies. To maintain their competitiveness in dynamic and turbulent environments, it is important for companies to anticipate and address these changes [17]. Otherwise, these turbulences will lead to a range of problems for companies such as reduced levels of service or higher inventory costs [46]. Scientists and practitioners alike argue that the respective technological changes will lead to the fourth industrial revolution [8, 10, 40, 54]. According to Broy, nothing has changed our lives as much over the last 40 years as the digital revolution [15]. Ongoing digitalization will enable firms to connect machines, storages and operating materials along their entire value chain. These underlying systems are called Cyber Physical Systems (CPS). In this context, the term 'Industrie 4.0' has become popular in Germany [37]. The core element of Industrie 4.0 is the vision of the smart factory that enables the use of the internet's intelligence for planning and executing production and increasing the agility of production systems [10, 40]. The human element is embedded in these systems as an actor (managing tasks), a problem solver and a collaborator [11, 33, 37].

P. Letmathe · M. Schinner (✉)
Chair of Management Accounting, RWTH Aachen University, Templergraben 64, 52062 Aachen, Germany
e-mail: Schinner@controlling.rwth-aachen.de

P. Letmathe
e-mail: Letmathe@controlling.rwth-aachen.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_25

It is obvious that the related technologies will lead to significant economic and social changes [27]. CPS can help to address key challenges related to an aging workforce or to scarce resources, but they also create a new form of complexity for the manufacturing industry [40]. Consequently, enterprises have to deal with this growing complexity and with the requirements of faster innovations and flexibility. One key factor for meeting these challenges will be to invest in employee competencies [68].

To fully utilize the potential of digitalization, companies will have to find the right balance between technology and human factors [22]. The traditional world of work has not prepared the workforce for the often demanding tasks of CPS. To achieve a good match with the technological challenges and complexity, firms need more long-term investments in the workforce (e.g. hiring and qualification) [24, 45]. Kölmel et al. [43] distinguish between technological complexity and contextual complexity, both rising due to the introduction of CPS. Technological complexity either refers to the fundamental interaction characteristics (input and output) of a technology, or to the fact that the underlying system architecture is complex, linking a variety of different systems, architectures, agents, databases, or devices. Contextual complexity includes the broader tasks, roles, or jobs that the technology is supposed to support, especially when tasks are open-ended or unstructured [43]. To handle these complexities, the development of technical as well as contextual competencies is crucial for the interaction between humans and technological systems [2]. Baxter and Sommerville [9] highlight that the failure to incorporate socio-technical approaches, which take necessary human competencies into account, tends to result in ill-defined requirements as well as poor system design, system delays and unmatched expectations [9]. Consequently, due to demographic factors (e.g. an aging workforce) and the more demanding skill requirements of CPS, companies need to maintain and develop the competencies of their workforce [60] in the areas of technological knowledge and contextual complexity.

To synchronize organizational and individual skills, it is important to analyze the required competencies for CPS and the actual individual competencies of the employees. The development of such competencies is time-consuming and costly. Hence, workforce management approaches should focus on the early identification and evaluation of competencies, which are relevant for the enterprise's strategy and for dealing with technological change. The identification of competence gaps and the planning of workforce requirements (hiring and qualification of employees) will become even more important in the future [10, 53]. Traditionally, such competencies are often clustered into technical, methodological, social, and personal competencies [34, 41, 53]. Technical and methodological competencies will play an important role in handling technological complexity. Furthermore, social and personal competencies are crucial to handle the contextual complexity of CPS. As employees will not be able to solve all problems individually, collaborative problem-solving capabilities will be even more important [22, 71].

In our article, we categorize different types of influences on these critical competencies in the future to master the technological and contextual complexity of CPS.

Furthermore, we introduce a short measurement instrument for these competencies that also includes aspects of technological and contextual complexity.

2 Cyber Physical Systems

According to Lee [48], "*Cyber-Physical Systems (CPS) are integrations of computation and physical processes. Embedded computers and networks monitor and control the physical processes, usually with feedback loops where physical processes affect computations and* vice versa" ([48], 1). CPS include embedded systems and devices. These could be, for example buildings, transport and medical devices as well as logistics, coordination and management processes or web services. Sensor systems collect data from physical systems and actors. Based on the evaluation and storage of data, CPS act or respond to the physical world with which they are connected locally or globally. Furthermore, they use data that are available worldwide and they have some multimodal human-computer interfaces [30]. Figure 1 illustrates a typical CPS architecture.

Applications of CPS have enormous potential. They can be applied in medical devices and systems, for traffic control and safety, in automotive systems, for process control, for energy conservation, in environmental control, in avionics, for

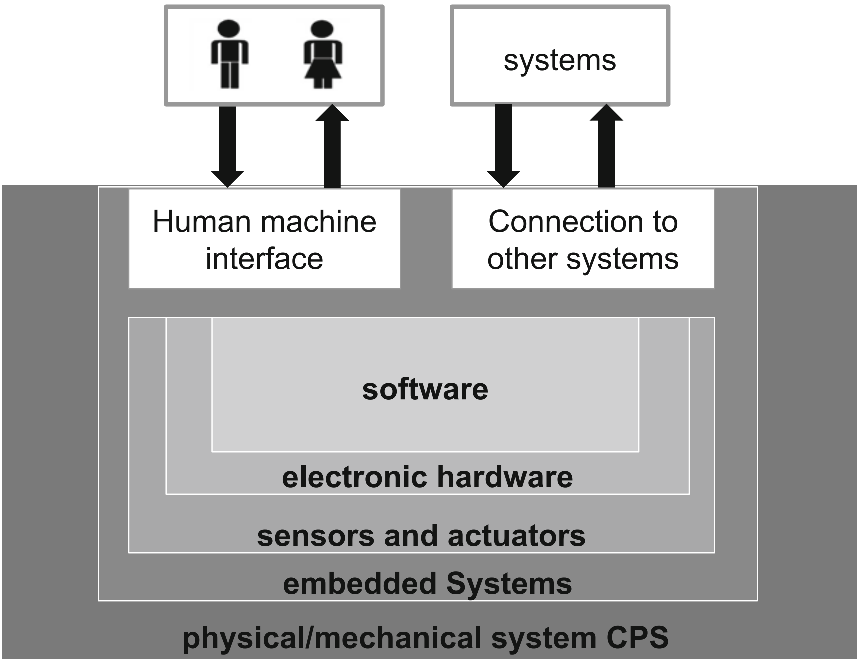


Fig. 1 CPS architecture [14, 15]

the control of critical infrastructures (e.g. electric power), in defense systems and in manufacturing, etc. For example, they can improve the efficiency and safety in transportation systems by connecting cars and smart traffic data systems, resulting in reduced fossil fuels consumption and lower greenhouse gas emissions [47].

With the objective of improving the productivity, the quality and the stability of production systems, manufacturing can be positively influenced by so called Cyber Physical Production Systems (CPPS) [38, 54]. Therefore, the introduction of CPS is a widely discussed topic, because CPS are expected to change business models and entire market structures [2]. According to Schlick et al. [65], the main characteristics of the change in production will be: smart objects, comprehensive networking, the use of internet standards for communication, adaptive and agile production systems, vertical integration in the network and the changed role of employees. The technological changes generate new opportunities for collaborative value creation, such as the potential to utilize the customer knowledge for the development and production processes [20, 72]. Owing to the opportunity to exchange information worldwide, labor becomes more independent from the locations of the manufacturing sites [44].

CPS has to be resilient and adaptable to unpredictable and also adverse events. Not every component is reliable [47]. The design and use of CPS is a considerable challenge involving specific requirements such as safety, usability or trust in the system [43, 48]. The introduction of CPS does not just concern technological change: it will also alter the role of employees, their collaboration and workplaces. Especially the human-machine interaction will require substantially different competence profiles of many employees. To address all these changes, the working group Industrie 4.0 has defined key priority areas with a need for business action or industrial policy for introducing CPS [40]:

- Standardization and open standards for a reference architecture: Information will be exchanged within and between companies. For a collaborative partnership, it is necessary to develop common standards. Furthermore, a reference architecture with a technical description would be helpful for implementing these standards.
- Managing complex systems: Trends, such as product customization, market requirements and increasing functionalities in complex production networks, are increasing production complexity. Explanatory and planning models can help to manage this complexity.
- Broadband infrastructure for industry: Existing communications networks have to be extended to meet the requirements of higher volumes and a better quality of data exchange.
- Safety and security as critical factors for the success of Industrie 4.0: The planned data and knowledge, which should be exchanged between companies or manufacturing facilities, have to be protected (security). Furthermore, the CPS (machines, products etc.) should not be a danger to employees or to customers (safety). Hence, CPS needs to be protected against misuse and unauthorized access.

- Work organization and work design in the digital industrial age: Industrie 4.0 will change the content of work processes and the role of employees. One reason for this is real-time-oriented control. Employees will be more involved with ad hoc problem solving instead of routine tasks. They will be more responsible for ensuring CPS stability and maintenance than for performing object-oriented tasks such as working on a part or a product. Furthermore, work design will be more participative and there will be a need for lifelong learning.
- Training and professional development for Industrie 4.0: Job and skill profiles will change through modified technological, social and organizational contexts. Current standardized training programs are limited and it will be a challenge to identify the relevant training contents. It is likely that interdisciplinary orientation and new qualifications will gain tremendous importance.
- Regulatory framework: The new complexity of digitalization cannot be mastered through existing regulatory frameworks. Modified regulatory frameworks will have to fulfill certain requirements, such as data protection, and should have the flexibility to utilize the potential benefits of new and rapidly changing technologies.
- Resource efficiency: One of the major goals of CPS implementation is to increase the productivity of production systems, i.e. to increase the ratio of output and resources, such as raw materials, energy, human and financial resources.

3 Competencies and Competence Management

This section discusses competence management as the basis of managing employees in the digitalized world. We distinguish between organizational and individual competencies and provide a short example of how the respective competencies can be classified and measured.

3.1 Defining Individual and Organizational Competencies

Technological inventions in the context of CPS are changing the avenues to competitive success and require the effective management of knowledge and employee skills [64]. But there are more than just knowledge and skills involved. McClelland [51] showed that conventional knowledge or ability tests cannot predict whether people can cope with the tasks of their jobs and he argues in favor of a behavioral and task-oriented analysis of competencies [51]. Competencies cannot be documented by certificates. Competencies are always related by actions in different situations [41]. However, there is no universally accepted definition of competencies. Reinhardt and North ([63], 1374) argue that *“A person’s competence* *basically describes a relation between requirements placed on a person/group or self-created requirements and these persons' skills and potentials to be able to meet these requirements. Competencies are concretized at the moment knowledge is applied and become measurable in the achieved result of the actions."* They regard competencies as being embodied in applied knowledge and measurable from the results of given tasks. This conforms with Sanchez ([64], 7), who stated that "*skills are the abilities an individual has to do things. Competency is the set of skills that an individual can use in doing a given task*". According to Erpenbeck and von Rosenstiel ([26], XIX), competencies are "*dispositions for self-organization activities*". Hence, competencies relate to problem-solving abilities, whereas it is possible to test qualifications in exams with always the same requirements. While test results reflect actual knowledge, they do not demonstrate whether somebody has the ability to transform this knowledge into sufficient action. In this sense, qualifications are dispositions of knowledge and skills [26].

The sum of all skills or abilities which an individual has and can use to fulfill tasks is the so called 'competence portfolio' of an individual. According to Reinhardt and North [63] the competence portfolio of an individual and that of an organization should be synchronized. Furthermore, there is a divided view of competence management. On the one hand, competence management is seen as a part of organizational science [57]. On the other hand, competence management is regarded as belonging to cognitive science [25, 63].

In the context of organizational competencies, it is possible to distinguish between capabilities and competencies. Prahalad and Hamel [61] summarize competencies as'the knowledge of a firm'. According to them, knowledge consists, for example of the skills in production or in technologies, or a combination of both. In contrast, capabilities are based more on (cross-functional) processes or routines within the business. Both competencies and capabilities are strategically relevant resources [50]. Organizational capabilities and competencies can be an important competitive advantage if they are costly, rare and valuable resources [4].

North et al. [58] argue that competence management should be aligned with the technology, the processes and the information technology infrastructure. Hence, it is important to differentiate between organizational competencies and individual competencies and to define those individual competencies that employees have to acquire to fulfill organizational requirements.

Reinhardt and North [63] call this 'competence adaptation'. The development of individual competencies should focus on the most relevant competencies/capabilities for successfully implementing the organization's routines and processes, such as knowledge in software engineering for embedded systems. Moreover, competence management should be aligned with the strategic and market decisions and the organizational structure. Reinhardt and North [63] already developed a model for matching organizational and individual perspectives. As discussed, companies have to achieve a fit between technologies—in this case the technology of CPS—and the competencies of humans. For this purpose, it is also important to describe and classify competencies from an organizational point of view in order to meet these requirements. Hafkesbrink and Schroll [35] describe organizational competencies for open innovation as the organizational readiness, the collaborative capability and the absorptive capacity. They argue that these organizational competencies can only be met by bundles of individual competencies enabling organizational members to collaboratively perform tasks with a low initial structure. Such collaborative competencies help to develop the absorptive capacities of the whole organization [35].

3.2 Classification and Measuring of Competencies

In the first part of this section, we discuss classifications of individual competencies. Then we go on to suggest a method for measuring competencies based on previous literature in this field.

3.2.1 Competence Classification

Stoof et al. [69] define five possible features of competence, which can be analyzed by answering the following five questions: 1. Is it a personal characteristic or is it a task-specific characteristic? 2. Is it the competence of an individual person or is it the competence of a team or an entire organization? 3. Is it a specific competence with a clearly defined scope that is not useful for other tasks, or is it a general competence with a broader scope within a profession or covering more than one profession? 4. Do different levels of a given competence exist or does a different level define a new competence? 5. Can the competence be taught, like knowledge, or can it not be taught?

Often competencies are divided into 'hard skills' (e.g. technical competencies) and 'soft skills' (e.g. communication competencies). There are already several classifications that have been developed for competencies to describe and distinguish them [26, 58]. According to Reinhardt and North [63] the portfolio of individual competencies can be divided into professional, methodological and social competencies. Crawford [18] elaborated a comprehensive approach to describe, categorize, identify and measure competence against standards for project management (see Fig. 2). She defines three categories: Input competencies, personal competencies and output competencies. Input competencies consist of knowledge and skills. Personal competencies are personality traits, attitudes and behaviors. Output competencies can be shown by demonstrative performance measures through a company's diagnostic systems [67].

Erpenbeck and von Rosenstiel [26] use four categories to describe competencies: methodological and professional competencies, personal competencies (willingness to learn), activity-and action-oriented competencies (e.g. flexibility) and socio-communicative competencies (e.g. team skills). We adopt a well-established approach [31, 34, 35, 41, 53] that distinguishes between functional (technical and—in parts—methodological) and cross-functional (social, self-management and—in

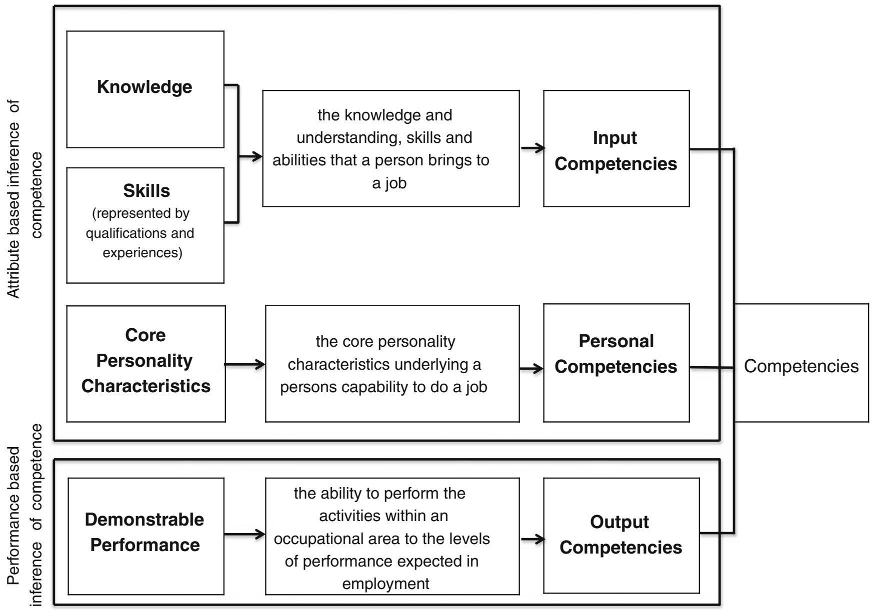


Fig. 2 Integrated model of competence identifying components of the overall construct [18]

parts—methodological skills) competencies. Technical competencies are knowledge, skills or experience, which are applicable in specific technical contexts. Methodological competencies encompass the application of teachable and well-defined methods, e.g. heuristic methods for solving complex problems. Social competencies reflect the ability to work successfully in teams and in cross-functional processes. Self-management competencies or personal competencies, on the other hand, help individuals to organize themselves efficiently, for example through self-control, self-organization and motivational competencies, such as the willingness to learn [31, 34, 41, 53] (see Fig. 3).

3.2.2 Competence Measurement

Although a widely accepted method for measuring competencies does not exist, several approaches for measuring competencies, e.g. with qualitative or quantitative analysis, have been suggested [26, 58]. These present a large overview of different competence measurement approaches with a focus on self-organization disposition from the German-speaking research community [26].

Learning from experience is particularly important for expanding necessary knowledge. For example, Barr et al. [5] and Pennings et al. [59] showed positive correlations between learning from experience and the success and renewal of a firm. Therefore, focusing on experience as an indicator of competence in a given

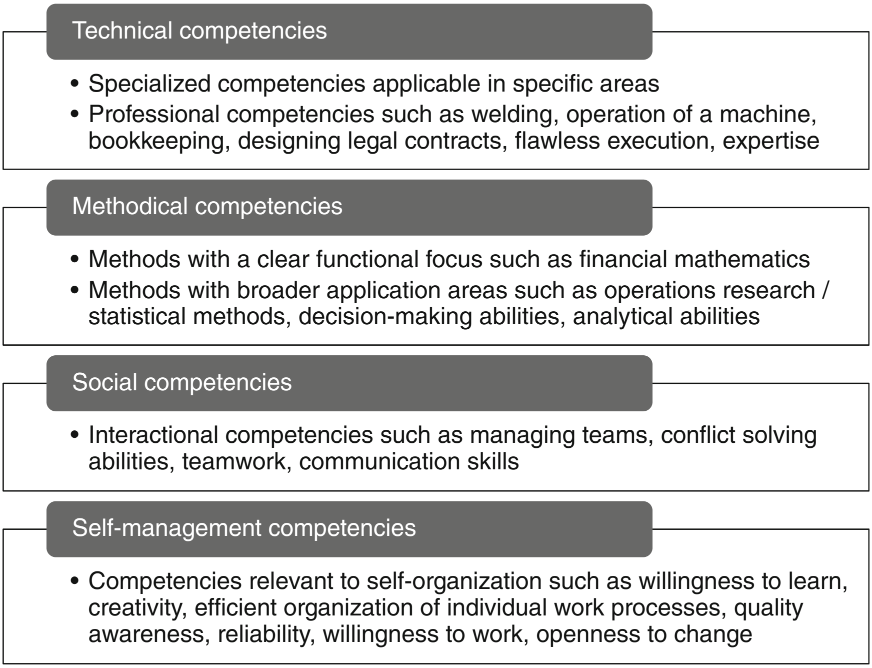


Fig. 3 Competence classification [34, 41, 53]

field can be beneficial for the evaluation of competencies in different dimensions. For the dimension of competencies, we adopt the proposal by North et al. [58], which can be adapted to different contexts. They propose four simple dimensions for the assessment of competencies: knowledge and experience, complexity of the task, autonomous working and self-management and reflection capability. These four dimensions are used to grade the competencies of an employee. Furthermore, they use a well-tested scale for the classification of three steps: connoisseur, experienced/advanced, and expert. Table 1 illustrates the described framework for the experience-based evaluation of competencies in their different dimensions.

The scale used relates to the European Language Portfolio proficiency levels (A1-C2) and is built on the four assessment dimensions shown above. These proficiency levels can be subdivided into six steps, comparable to the European Language Portfolio. By using this scale it is possible to attribute a qualitative disposition to a competence [58].

Table 1 Evaluation of Competencies (based on North et al. [58, 70–71])[a]

Dimensions	Connoisseur		Experienced and advanced		Expert (can train others)	
	A1	A2	B1	B2	C1	C2
Knowledge and experience	"What level of knowledge and what degree of experience are required in a specific field? Basic knowledge, differentiated areas of expertise or comprehensive knowledge with varied application experience?"					
Task complexity	"The degree of complexity depends on how many relevant factors exist and the mutual dependencies between these factors"					
Autonomous work and self-management	"The path, the goal and the willingness to fulfill the task are the three key aspects of this evaluation dimension"					
Capability of reflection	"Competence means always reflecting on your own actions or the actions of others in the context of a situation. To what extent am I capable of critically reflecting on processes, situations, people and behavior as to whether they meet the expected requirements?"					

[a]The original quotes have been translated from the German into English by the authors of this paper

4 Consequences and New Competence Requirements for Employees Through CPS Complexity

The consequences of introducing CPS are being widely discussed but are mostly based on speculation. One exception is the study by Frey and Osborne [29], which received a great deal of attention. According to Frey and Osborne [29], 47 % of the jobs in the US have a high risk of being automated over the next 10 to 20 years. Furthermore, the probability of automation declines in line with the level of salary and education of the employee. Bonin et al. [13] conducted a similar study for Germany on a more detailed level for different types of tasks. They found that only 12 % of current job tasks in Germany and 9 % of those in the US have a high probability of being automated over the course of the coming years. However, according to these studies, the nature of existing tasks will shift in focus and towards a higher degree of complexity. In this context Kölmel et al. [43] distinguish between technical complexity and contextual complexity (see Table 2). Technical complexity refers to the interaction characteristics and the architecture or databases of the system [32]. Contextual complexity particularly refers to a change in the nature of existing tasks. Tasks performed by humans will become more unstructured and employees will have to perform a wider range of tasks, with their roles

Table 2 Technical and contextual complexity of CPS task characteristics based on: Bonin et al. [13]; Dworschak and Zaiser [22]; Dworschak et al. [23]; Frey and Osborne [29])

	Technological complexity	Contextual complexity
Increasing challenges of CPS for the workforce	• Interaction characteristics technology (interfaces, coordination, information exchange, systems stability) • Systems architecture and variety of different systems, agents, architectures, devices, or databases	• Broader tasks, roles or jobs • Open-ended and unstructured tasks (problems) • Less structure • Abstractness • Interpretation and use of information • Collaboration • Information overload

changing towards problem solving and collaborative work. Competencies such as obtaining sufficient information and interpreting data correctly will become more important. Routine tasks with clearly defined steps and results will become increasingly more automated [13, 22, 29].

As a result of automation, it is not sufficient to train routine tasks and to develop all necessary competencies. Employees also have to make interventions if unexpected and complex non-routine problems occur [71]. Gorecky et al. [33] also see the primary task in setting and supervising the realization of the production strategy for a set of production facilities. In these views, humans are acting as creative problem solvers for complex problems or opening up new optimization potentials. According to McCreery and Krajewski [52], complex tasks are associated with slow learning and fast forgetting and simple tasks with fast learning and slow forgetting. Therefore, companies have to identify and develop the appropriate combination of competencies and in particular the methodological knowledge that is required to analyze and solve problems. Moreover, self-management competencies will become crucial for employee performance, whereby employees will have to learn how to obtain the necessary knowledge and information. Following this view of complexity, three factors will gain key importance for the successful implementation of CPS that will also influence the contextual complexity [71]:

- The role of employees
- The content of the task
- The interaction and collaboration of humans with systems and within work teams

Table 3 Scenarios in CPS [22]

Automation Scenario	Tool scenario
• CPS guides skilled workers • Work is determined by technology • Autonomy of skilled workers is limited • Emergence of a skill gap: Skilled workers cannot develop/build up the know-how for dealing with problems anymore • High-skilled employees are responsible for installation, modification and maintenance of CPS	• Skilled workers guide CPS • CPS is the central domain of skilled workers • CPS supports the decision-making of skilled workers • A successful performance requires the provision of crucial information and suitable approaches of vocational education and training due to an increasing demand for IT, electronic and mechanical knowledge

The role of employees: Windelband et al. [71] analyzed the future skill requirements in the field of logistics and developed two possible scenarios (see Table 3) within this study [22, 71]. Both scenarios show that the roles of employees will change substantially. Under the human-centered tool scenario, humans make the major decisions guided by the CPS and take corrective action in the automated process. Under the automation scenario, decision-making shifts to the technical sphere of the production system [23]. Furthermore, the employees with different interdisciplinary backgrounds have to coordinate themselves. Under the latter scenario, intelligent CPS can run the entire production and human competencies are only needed when systems are installed, modified and updated or if problems occur. In addition, employees will also be required as high-level problem solvers when machine intelligence is not able to deal with emergencies, system failure or other problems [10, 11, 22, 71].
The content of the task: Autor and Dorn [3] hypothesize that employees will perform more creative, problem-solving and coordination tasks that cannot be substituted by computers and algorithms. Frey and Osborne [29] refer to this as'engineering bottlenecks', where computers cannot substitute humans. These bottlenecks are tasks that require perception and manipulation, social intelligence or creativity [29].

- **Perception and manipulation tasks**: For these tasks, employees require skills that enable them to structure and understand *complex and unstructured environments*. For example, machines are often not able to identify complex process failures and to develop solutions for some problems, such as finding a mislabeled inventory item and reentering it into the process [13, 29].

- **Creativity tasks**: Creativity is the ability to develop ideas or artifacts, which are new and valuable [12]. These ideas could be for instance poems, cooking recipes or scientific theories or artifacts. Creativity is driven by using the brain's associative platforms and pattern recognition and is supported by the brain's complex network of neurons. Therefore, it cannot be expected that CPS will be able to fully substitute human creativity over the course of the next decades [13, 29].
- **Collaborative tasks**: Social competencies can be regarded as the lubrication oil of organizations, ensuring that collaboration and processes yield the desired results. Typical examples of collaborative tasks that heavily rely on social intelligence are negotiations between two partners or the motivation of employees. Admittedly, some computers can already imitate the social interactions between humans with algorithms, but there is still the factor of human emotions during interaction processes and computers are not (yet) able to master the complexity of human interaction processes [29]. Frazon et al. [28] highlight flexibility and in particular the problem-solving competence of humans to develop their full potential. These soft skill aspects are often neglected in CPS research, and only a few approaches have attempted to analyze it [70]. Solving complex and unstructured problems requires more cognitive operations and thus more cognitive skills [39] compared to solving non-complex and well-structured problems [42]. As a result, we also have to emphasize altered interaction and collaboration, which consequently requires increased social intelligence. Table 4 summarizes the changes of the task content.

The interaction and collaboration of humans with systems and within work teams: CPS will change the interaction with and the control of the physical world by humans [62]. The interaction between humans and machines through sensors and also different interactions between humans will be an enormous challenge and might even question the acceptance of CPS by humans [2]. The traditional workplace in the office will become less important, because digital networks and the availability of real-time data allow physical production activities to be managed

Table 4 Task content scenarios (based on: Bonin et al. [13]; Dworschak and [22]; Dworschak et al. [23]; Frey and Osborne [29])

	Traditional Industry	CPS and Industrie 4.0
Task content	• More routine tasks • Focused on one discipline • More structured tasks with a clear goal	• Fewer routine tasks • More interdisciplinary problems • Unstructured tasks • Perception and manipulation tasks • Collaborative tasks • Creative tasks • Flexibility • Increased scope for decision making

from anywhere. As a result, the task spectrum of many employees will increase [33]. In addition, CPS will increasingly involve diverse groups in communication and interaction processes. Consequently, there will be a need for new concepts of collaboration [49] and structuring of the work between humans and machines [10]. Schuh et al. [66] present a framework for collaborative practices in Industrie 4.0 environments and they exemplify that different dimensions of collaboration (communication, coordination, cooperation) can be levers for CPS. Hirsch-Kreinsen [36] and others highlight that employees with high qualifications and high flexibility within a loose network will have to solve problems collaboratively and in a self-organized manner—a concept comparable to swarm intelligence. They will use informal social processes for communication and cooperation to organize their specialized knowledge [19, 36, 56]. Concepts such as open production or open innovation will become more relevant [6] and foster these trends and the need for a closer look at the necessary competence categories. Especially communication and interaction (human-to-human and human-to-machine) are vital aspects of these concepts. Even though technical competencies will remain important, it is obvious that soft skills (social, personaland—and in parts—also methodological skills) will be given increased attention in relation to professional competence [7, 53, 55].

5 Development of a Measurement Instrument for Competencies in the Age of CPS

To categorize and to measure the necessary competencies for CPS, we propose the following scale that has been derived and further developed from that of North et al. [58] and adapted to the requirements determined for CPS. Table 5 summarizes our discussion. In general, we conclude that CPS will stimulate a shift of many of the criteria to the right-hand side of Table 5, indicating less structured tasks and more interdisciplinary collaboration and problem-solving abilities. As a result, we divide the already existing dimension complexity of the tasks into technological complexity and contextual complexity. For contextual complexity, we distinguish three different dimensions of task structure, task content and interactiveness. Self-management skills and reflection capabilities complete our taxonomy.

Table 5 Evaluation of competencies for CPS (based and exentended on: North et al. [58])

		Connoisseur		Experienced and advanced		Expert and creative problem solver	
Role		▪ Operator with low or basic competence levels		▪ Experienced operator with intermediate competence levels		▪ Creative problem solver ▪ Decision maker ▪ Teacher	
		A1	A2	B1	B2	C1	C2
Knowledge and experience		Basic Knowledge		⟷		Detailed knowledge and broad experience from different contexts and ability to train others	
Technological complexity of the task		No technical background necessary		⟷		Challenging technical complexity with new and different systems and intensive interaction in interdisciplinary teams	
Contextual complexity of the task	Structure of the task	Clearly structured tasks		⟷		Challenging and new, unstructured tasks changing, depending on different and unknown contexts	
	Content of the task	Routine tasks		⟷		Previously unknown situations and tasks, in interdisciplinary teams, creative solutions are required	
	Inter-action and collabo-ration	Single discipline without any technical interaction and collaboration with others		⟷		High interdisciplinary and rapidly changing teams with different backgrounds; interaction only through technical interfaces with human and machine intelligence	
Independent work and self-management		Work under guidance and with support from others		⟷		Independent and flexible work requiring creative and innovative solutions Interaction with machines and humans Leadership skills are crucial	
Reflection capability		I can judge my actions and optimize them within the given framework		⟷		I can reflect on my actions, detect errors and misconduct and can use my knowledge for the expansion, differentiation and optimization of my actions	

6 Conclusions

The manufacturing environment is changing, as is the environment for employees in the manufacturing industry. Increasingly, routine jobs or tasks will be automated. Employees have to handle new technological and contextual complexities which determine their new role, content of tasks as well as interaction and collaboration procedures. Technological complexities can be addressed by refining the content of traditional qualifications. A rising contextual complexity is becoming relevant, however, especially in tasks which require more social and personal competencies. Due to factors such as increased flexibility, the roles of employees will change substantially. Employees will have to broaden their competencies in order to handle unstructured situations involving uncertainty. Furthermore, it will be necessary for them to work in teams with different interdisciplinary backgrounds in order to solve problems. Often, it will be necessary to communicate via interfaces, in different languages and across different time zones.

We have categorized different types of competencies which will be necessary for employees to work successfully in CPS environments. There is an increasing challenge for employees to learn continuously and to develop on-the-job competencies. Therefore, we suggest a measurement instrument for these competencies and demonstrate how the included technical and collaborative competencies provide guidance for mastering the technological and contextual complexity of CPS. Planning and managing these critical competencies are a crucial factor of CPS-based production systems. With the rise of new technologies there is also a need to define organizational competencies more precisely in order to successfully handle these complex environments.

These new roles, task content and interaction behaviors might overburden employees however, while simple and repetitive tasks will become increasingly automated [36]. The success of workers will depend on their flexibility, problem solving competencies as well as their willingness to engage in lifelong learning; otherwise, they will not be able to keep up with the required changes in their workplaces and work procedures. This challenge might also explain why many companies are reluctant to invest in CPS. We therefore conclude that competence management on the organizational level as well as the reform of public education (including the German apprenticeship/ trainee system) are important factors for introducing CPS.

References

1. Abele E, Reinhart G (2011) Zukunft der Produktion: Herausforderungen, Forschungsfelder, Chancen. Carl Hanser Verlag, München
2. Acatech (2011) Cyber physical systems. Driving force for innovation in mobility, health, energy and production. Acatech, Berlin
3. Autor DH, Dorn D (2013) The growth of low skill service jobs and the polarization of the US labour market. Am Econ Rev 103(5):1553–1597

4. Barney JB (1995) Looking inside for competitive advantage. Acad Manag Executive 9(4):49–61
5. Barr PS, Stimpert JL, Huff AS (1992) Cognitive change, strategic action, and organizational renewal. Strateg Manag J 13(S1):15–36
6. Basmer S, Buxbaum-Conradi S, Krenz P et al (2015) Open production: chances for social sustainability in manufacturing. Procedia CIRP 26:46–51
7. Bauer W, Hämmerle M, Schlund S et al (2015) Transforming to a hyper-connected society and economy—towards an "Industry 4.0". Procedia Manuf 3:417–424
8. Bauernhansl T (2014) Die Vierte Industrielle Revolution—DerWeg in ein wertschaffendes Produktionsparadigma. In: Bauernhansl T, ten Hompel M, Vogel-Heuser B (eds) Industrie 4.0 in Produktion, Automatisierung und Logistik—Anwendung—Technologien—Migration. Springer, Wiesbaden, pp 5–35
9. Baxter G, Sommerville I (2011) Socio-technical systems: from design methods to systems engineering. Interact Comput 23(1):4–17
10. Becker KD (2014) Arbeit in der Industrie 4.0—Erwartungen des Instituts für angewandte Arbeitswissenschaft e.V.. In: Botthof A, Hartmann EA (eds) Zukunft der Industrie 4.0. Springer, Berlin, pp 23–30
11. Bochum U (2014) Gewerkschaftliche Positionen in Bezug auf „Industrie 4.0“. In: Botthof A, Hartmann EA (eds) Zukunft der Industrie 4.0. Springer, Berlin, pp 31–46
12. Boden MA (2003) The creative mind: myths and mechanisms. Routledge, London
13. Bonin H, Gregory T, Zierahn U (2015) Endbericht—Kurzexpertise Nr. 57—Übertragung der Studie von Frey/Osborne (2013) auf Deutschland. Bundesministerium für Arbeit und Soziales. ftp://zew.de/pub/zew-docs/gutachten/Kurzexpertise_BMAS_ZEW2015.pdf. Accessed 10 Apr 2016
14. Brettel M, Friederichsen N, Keller M et al (2014) How virtualization, decentralization and network building change the manufacturing landscape: an Industry 4.0 perspective. Int J Mech Aerosp Ind Mechatron Manuf Eng 8(1):37–44
15. Broy M (2010) Cyber-Change—WissenschaftlicheHerausforderungen bei der Entwicklung. In: Broy M (ed) Cyber Physical Systems: Innovation durch softwareintensive eingebette Systeme, acatech DISKUTIERT Springer, Berlin, pp 17–32
16. Calo TJ (2008) Talent management in the era of the aging workforce: the critical role of knowledge transfer. Public Pers Manag 37(4):403–416
17. Cao M, Zang Q (2008) A path analytical analysis of organizational integration practices' effects on manufacturing flexibility and competitive advantage. J Syst Sci Syst Eng 17 (3):277–296
18. Crawford L (2005) Senior management perceptions of project management competence. Int J Project Manag 23(1):7–16
19. Cummings M, Bruni S (2009) Collaborative human-automation decision making. In: Nof SY (ed) Springer handbook of automation. Springer, Berlin, pp 437–447
20. De Grave A, Olsen SI, Hansen HN et al (2010) Sustainability of micro-manufacturing technologies. In: Qin Y (ed) Micromanufacturing engineering and technology. William Andrew Elsevier, Oxford, pp 394–404
21. De Long DW (2004) Lost knowledge: confronting the threat of an aging workforce. Oxford University Press, New York
22. Dworschak B, Zaiser H (2014) Competences for cyber physical systems in manufacturing—first findings and scenarios. Procedia CIRP 25:345–350
23. Dworschak B, Zaiser H, Martinez S et al (2013) Logistics as a domain of application for the "Internet of Things" in the early identification initiative of the German Federal Ministry of Education and Research (BMBF). In: The future skills needed for the Internet of Things in the field of logistics, Frequenz—Summary of study findings. http://www.frequenz.net/uploads/tx_freqprojerg/EN_Summary_logistics_final.pdf. Accessed 10 Apr 2016
24. Egan MJ (2007) Anticipating future vulnerability: defining characteristics of increasingly critical infrastructure-like systems. J Conting Crisis Manag 15(1):4–17

25. Erpenbeck J, Heyse V (2007) Die Kompetenzbiographie: Wege der Kompetenzentwicklung. Waxmann, Münster
26. Erpenbeck J, von Rosenstiel L (2007) Einführung. In: Erpenbeck J, von Rosenstiel L (eds) Handbuch Kompetenzmessung: Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis. Schaeffer-Poeschel, Stuttgart, pp. XVII-XLVI
27. Evangelista R, Guerrieri P, Meliciani V (2014) The economic impact of digital technologies in Europe. Econ Innov New Technol 23(8):802–824
28. Frazon EM, Hartmann J, Makuschewitz T et al (2013) Towards socio-cyber-physical systems in production networks. Procedia CIRP 7:49–54
29. Frey C, Osborne M A (2013) The future of employment: how susceptible are jobs to computerization? University of Oxford, pp 1–72
30. Geisberger E, Broy M (2012) agenda CPS – Integrierte Forschungsagenda Cyber-Physical Systems. Springer, Berlin
31. Gerst D (2015) Designing workplaces from a work organizational perspective. In: Wiendahl HP, Reichardt J, Nyhuis P (eds) Handbook factory planning and design. Springer, Heidelberg, pp 169–195
32. Golightly D, D'Cruz M, Patel H et al (2011) Novel interaction styles, complex working contexts, and the role of usability. In: Albers MJ, Still B (eds) Usability of complex information systems evaluation of user interaction. CRC Press, Boca Raton, pp 281–304
33. Gorecky D, Schmitt M, Loskyll M (2014): Mensch-Maschine-Interaktion im Industrie 4.0-Zeitalter. In: Bauernhansl T, ten Hompel M, Vogel-Heuser B (eds) Industrie 4.0 in Produktion, Automatisierung und Logistik—Anwendung—Technologien—Migration. Springer Wiesbaden, pp 525–527
34. Grote S, Kauffeld S, Billich M et al (2006) Implementierung eines Kompetenz-managementsystems: Phasen, Vorgehen und Stolpersteine. In: Grote S, Kauffeld S, Frieling E (eds) Kompetenzmanagement—Grundlagen und Praxisbeispiele. Schaeffer-Poeschel, Stuttgart, pp 33–58
35. Hafkesbrink J, Schroll M (2010) Organizational competences for open innovation in small and medium sized enterprises of the digital economy. In: Hafkesbrink J, Hoppe HU, Schlichter J (eds) Competence management for open innovation. EUL Verlag, Lohmar Köln, pp 21–52
36. Hirsch-Kreinsen H (2014) Entwicklungsperspektiven von Produktionsarbeit. In: Botthof A, Hartmann EA (eds) Zukunft der Arbeit in Industrie 4.0. Springer, Berlin, pp 89–98
37. Hirsch-Kreinsen H (2015) Einleitung: Digitalisierung industrieller Arbeit. In: Hirsch-Kreinsen H, Ittermann P, Niehaus J (eds) Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen. Nomos, Baden-Baden, pp 10–31
38. Ittermann P, Niehaus J (2015) Industrie 4.0 und Wandel von Industriearbeit—Überblick über Forschungsstand und Trendbestimmungen. In: Hirsch-Kreinsen H, Ittermann P, Niehaus J (eds) Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen. Nomos, Baden-Baden, pp 33–52
39. Jonassen DH (2000) Toward a design theory of problem solving. Educ Tech Res Dev 48 (4):63–85
40. Kagermann H, Wahlster W, Helbig J (2013) Securing the future of German manufacturing industry—recommendations for implementing the strategic initiative INDUSTRIE 4.0—Final report of the Industrie 4.0 Working. Plattform Industrie 4.0, Frankfurt/Main.http://www.acatech.de/fileadmin/user_upload/Baumstruktur_nach_Website/Acatech/root/de/Material_fuer_Sonderseiten/Industrie_4.0/Final_report__Industrie_4.0_accessible.pdf. Accessed 10 Apr 2016
41. Kauffeld S (2006) Kompetenzen messen, bewerten, entwickeln: Ein prozessanalytischer Ansatz für Gruppen. Schäffer-Poeschel Verlag, Stuttgart
42. Kluwe RH (1995) Single case studies and models of complex problem solving. In: Frensch PA, Funke J (eds) Complex problem solving: the European perspective. Psychology Press, Hillsdale, pp 269–291

43. Kölmel B, Bulander R, Dittmann U et al (2014) Usability requirements for complex cyber physical systems in a totally networked world. In: Camarinha-Matos LM, Afsarmanesh H (eds) Advances in production management systems. Sustainable production and service supply chains—15th IFIP WG 5.5 working conference on virtual enterprises, PRO-VE 2014, Amsterdam, The Netherlands, 6–8October 2014, vol. 2. Springer, Heidelberg, pp 253–258
44. Krenz P, Wulfsberg JP, Bruhns FL (2012) Unfold Collective Intelligence. Zeitschrift für wirtschaftlichen Fabrikbetrieb 3:152–157
45. Kroes P, Franssen M, Poel I et al (2006) Treating socio-technical systems as engineering systems: some conceptual problems. Systems Research and Behavioural Science 23(6):804–814
46. Kulp S, Lee H, Ofek E (2004) Manufacturer benefits from information integration with retail customers. Manage Sci 50(4):431–444
47. Lee EA (2007) Computing foundations and practice for cyber physical systems: a preliminary report. Technical report no. UCB/EECS-2007–72:1-25
48. Lee EA (2008) Cyber physical systems: design challenges. Technical report no. UCB/EECS-2008-8, pp 1-8
49. Linke A (2015) Online-Kommunikation in einem sich verändernden Kommunikations-umfeld. In: Linke A (ed) Management der Online-Kommunikation von Unternehmen. Springer Fachmedien, Wiesbaden, pp 99–132
50. Marino EM (1996) Developing consensus on firm competencies and capabilities. Acad Manag Executive 10(3):40–51
51. McClelland DC (1973) Testing for competence rather than for "Intelligence". Am Psychol 28 (1):1–14
52. McCreery JK, Krajewski LJ (1999) Improving performance using workforce flexibility in an assembly environment with learning and forgetting effects. Int J Prod Res 37(9):2031–2058
53. Meyer G, Brüning B, Nyhuis P (2015) Employee competences in manufacturing companies —an expert survey. J Manag Dev 34(8):1004–1018
54. Monostori L (2014) Cyber-physical production systems: roots, expectations and R&D challenges. Procedia CIRP 17:9–13
55. Moraal D, Lorig B, Schreiber D et al (2009) BIBB Report Forschungs- und Arbeitsergebnisse aus dem Bundesinstitut für Berufsbildung—Ein Blick hinter die Kulissen der betrieblichen Weiterbildung in Deutschland—Daten und Fakten der nationalen CVTS3-Zusatzerhebung. BIBB Rep 2(7):1–12
56. Neef A, Burmeister K (2005) Die Schwarm-Organisation—Einneues Paradigma für das e-Unternehmen der Zukunft. In: Kuglin B, Thielmann H (eds) Real-Time Enterprise in der Praxis. Springer, Berlin Heidelberg, pp 563–572
57. North K (2011) Wissensorientierte Unternehmensführung—Wertschöpfung durch Wissen. Springer Gabler, Wiesbaden
58. North K, Reinhardt K, Sieber-Suter B (2013) Kompetenzmanagement in der Praxis: Mitarbeiterkompetenzen systematisch identifizieren, nutzen und entwickeln. Mit vielen Fallbeispielen. Springer Gabler, Wiesbaden
59. Pennings JM, Barkema HG, Douma SW (1994) Organizational learning and diversification. Acad Manag J 37(3):608–640
60. Plattform Industrie 4.0 (2015): Industrie 4.0—Whitepaper FuE-Themen
61. Prahalad CK, Hamel G (1990) The core competence of the corporation. Harv Bus Rev 68:79–91
62. Rajkumar R, Lee I, Sha L et al (2010) Cyber physical systems: the next computing revolution. In: Design automation conference 2010, Anaheim, pp 731–736
63. Reinhardt K, North K (2003) Transparency and transfer of individual competencies—a concept of integrative competence management. J Univ Comput Sci 9(12):1372–1380
64. Sanchez R (2001) Managing knowledge into competence: the five learning cycles of the competent organization. In: Sanchez R (ed) Knowledge management and organizational competence. Oxford University Press, Oxford, pp 3–38

65. Schlick J, Stephan P, Zühlke D (2012) Produktion 2020: Auf dem Weg zur 4. Industriellen Revolution. IM—Fachzeitschrift Information Management und Consulting 27(3):26–33
66. Schuh G, Potente T, Varandani R et al (2014) Collaboration moves productivity to the next level. Procedia CIRP 17:3–8
67. Simons R (2000) Performance measurement and control systems for implementing strategy: text and cases. Prentice Hall, Upper Saddle River
68. Spath D, Ganschar O, Gerlach S et al (2013) Produktionsarbeit der Zukunft—Industrie 4.0. Fraunhofer Verlag, Stuttgart
69. Stoof A, Martens RL, Van Merrienboer JJG et al (2002) The boundary approach of competence: a constructivist aid for understanding and using the concept of competence. Hum Resour Dev Rev 1(3):345–365
70. Wang W (2010) The emergence of intelligent enterprises: from CPS to CPSS. IEEE Intell Syst 25(4):5–88
71. Windelband L, Fenzl C, Hunecker F et al (2013) Future skill requirements engendered by the "Internet of Things" in the field of logistics. In: The future skills needed for the Internet of Things in the field of logistics, Frequenz—Summary of study findings. http://www.frequenz.net/uploads/tx_freqprojerg/EN_Summary_logistics_final.pdf. Accessed 10 Apr 2016
72. Wulfsberg JP, Redlich T, Bruhns FL (2011) Open production: scientific foundation for co-creative product realization. Prod Eng Res Devel 5(2):127–139

Part VII
Adjacent Fields and Ecosystems

Cyber-Physical Systems for Agricultural and Construction Machinery—Current Applications and Future Potential

Georg Jacobs, Felix Schlüter, Jan Schröter, Achim Feldermann and Felix Strassburger

1 Introduction

In the past, the developments in the agricultural and construction sector have experienced major jumps at certain times. These times are referred to as industrial revolutions and each of them is based on a different invention or technology. The timeline as well as the enabling technologies for the four industrial revolutions in agriculture are shown in Fig. 1. In the construction sector, the development has been very similar with a varying timeline.

Since the third industrial revolution, the number and quality of sensors as well as electronically controlled actuators have increased tremendously. New connection technologies, wire-based as well as wireless, enable fast data transfer over short and long distances. Furthermore, increased computing capacity allows much faster and more complex processing of the collected data. The combination of all these enablers leads to Cyber-Physical Systems, dense and strong connections between cyber entities and the physical world. These CPS allow a higher use of robotic and automatic support for certain work tasks and are currently of high interest for future

G. Jacobs (✉) · F. Schlüter · J. Schröter · A. Feldermann · F. Strassburger (✉)
Institute for Machine Elements and Machine Design (IME),
RWTH Aachen University, Schinkelstr. 10, 52062 Aachen, Germany
e-mail: jacobs@ime.rwth-aachen.de

F. Strassburger
e-mail: strassburger@ime.rwth-aachen.de

F. Schlüter
e-mail: schlueter@ime.rwth-aachen.de

J. Schröter
e-mail: schroeter@ime.rwth-aachen.de

A. Feldermann
e-mail: feldermann@ime.rwth-aachen.de

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_26

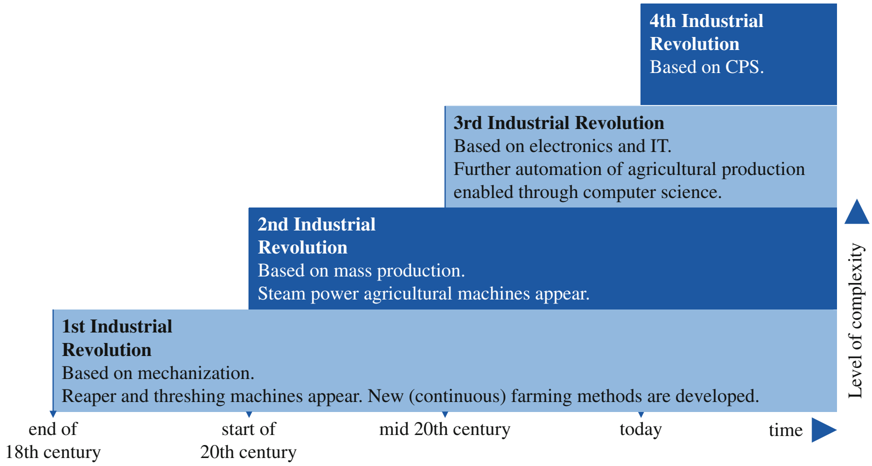


Fig. 1 From 1st to 4th industrial revolution in agriculture

developments, showing up in great numbers of patent applications [9]. Also, CPS can help create integrated and self-regulating systems across boundaries of single companies or industries as they are an essential enabler for the optimization of processes [79]. In both sectors, agriculture and construction, summed up as mobile machinery sector below, Cyber-Physical Systems are also the enabling technology which will lead to the fourth industrial revolution [79]. Through these CPS, increased operating efficiency and enhanced process efficiency [26] as well as better safety and a reduced environmental impact [27] can be achieved.

This chapter presents an overview of current applications and future potentials of Cyber-Physical Systems in the agricultural and construction machinery sectors. At first, the different challenges in both sectors are explained and a typical CPS for mobile machines is described. This is followed by a categorization of data in mobile machinery and a description of the future impact on communication strategies. In Sect. 5, the key technologies enabling the creation of CPS are discussed in the context of agricultural and construction machinery. Thereafter, the key algorithms, which enable the definition of strategies by smartly and efficiently processing and combining data, are described. This chapter ends with a detailed analysis of one typical agricultural and one typical construction process, in which the applications of the key technologies and key algorithms are exemplified.

2 Challenges

As already described, the main motivation for CPS is increased operating and process efficiency [26] as well as better safety and a reduced environmental impact [27]. However, the detailed challenges in agricultural and construction machineries differ and are therefore described separately in this section.

2.1 Challenges in Agricultural Machinery

A major challenge for today's society is the growth of the world's population, which reached 7.2 billion in 2014. By 2050, the population is expected to increase by more than 2 billion [105]. The associated growing need for food production cannot be solved by expanding cultivated land—due to its limited availability—so that farming has to increase its yield. Of course, farming also has to be economical and the operating costs (seeds, fertilizer, fuel, labor force, machinery, etc.) have to decrease. Therefore, the size and speed of agricultural machines have continuously been increased in the past decades, today reaching legal limitations in weight or dimensions [56]. These high-capacity machines themselves do not work sustainably, however, since they treat fields or livestock uniformly despite their inconsistent properties. Within local regions, the properties of cultivated land can differ significantly. Farmers have optimized the yield by adapting to this for centuries. However, with limited available resources on the planet, both, increased yield and sustainable cultivation, can only be achieved by even more precise farming of the existing cultivated land and by using these high-capacity machines.

Differences also exist within fields, because characteristics like slopes, soil properties or water supply vary. Precise adaptation to these locally varying conditions is necessary, but it is not possible to achieve more precise farming solely through the farmer or by using individual technologies. Only a site-specific field management system or even a whole-farm management approach offer further improvements in yield and sustainability. Today, this approach is state of the art and termed as Precision Agriculture (PA). PA is based on Cyber-Physical Systems (see Sect. 4) since it consists of the characteristic elements of a CPS:

For the whole farm management system, *sensors* are needed to record the state of the *processed goods* (soil, crop, livestock), the agricultural *machinery* and the *environment*. The collected *data* is held available with a database, processed by *algorithms* (strategies) and transferred via *interfaces*. Based on the obtained information, *actuators* adjust the agricultural machines.

A study by the European Parliament "confirms that Precision Agriculture can play a substantial role … in meeting the increasing demand for food, feed, and raw materials, while ensuring sustainable use of natural resources and the environment" ([98], pp 9). However, there is still great potential for improvement. Individual PA technologies and PA systems are available and established in the market. Whole cyber-physical farm management systems, which crosslink these individual PA systems, exist—but their efficiency in practical use is yet to be proven. According to Griepentrog "farmers with the new technology find themselves drowning in a flood of data without knowing how to apply it" ([47], pp 24). And still more data—including data for off-farm matters such as logistics and economy—will be available and integrated into agricultural tasks, increasing their complexity even more. This leads to an increased research demand on CPS, focusing especially on the fusion of data [48] but also on developing and crosslinking new technologies (e.g. sensors and mobile machines).

2.2 Challenges in Construction Machinery

As construction sites become larger, more complex and—most critically—are planned much more precisely, the necessary strategies and tools encounter a similar transformation. In a construction process, the crosslinking of the single value-adding steps is much more important. This is due to the fact that multiple machines simultaneously work on one construction site while the exact specification for the task of one machine highly depends on the result of other machines. These single steps are planned at a very detailed level regarding geometrical requirements, timing and organizational responsibility, since any deviation from the planned result influences all other steps. In the end, the goal is not to optimize a specific step, but the process as a whole. Due to the simultaneous execution of work, a quick adaptation to the new circumstances is necessary to minimize an undesirable impact if process results deviate. This means that machines performing other tasks as well as the supervising system need to receive current information as soon as possible, ideally in real time. Furthermore, the optimal performance regarding a specific task is measured by many variables, for example time, fuel consumption or the number of machines and workers involved. This calls for a multi-criteria performance evaluation, which adapts to the actual needs of the whole construction process. Hence, evaluating the efficiency of smarter machinery will also require smarter evaluation methods [27]. In order to fulfill the goal of optimizing the overall process, a higher availability of data and an enhanced connectivity between the involved sensors, devices, machines and planning and control tools are required. In the described system, physical components (devices, actuators, materials) interact with cyber entities (data storage and processing, various planning tools), which makes this a Cyber-Physical System. Enhanced connectivity of existing elements, the addition of further elements to the system, greater and smarter evaluation of existing data as well as the development of new sensors delivering information unavailable today will lead to the further development of CPS. Its use results in higher automation in construction, enabling new levels of machinery performance [12]. Only this development can make it possible to reach the overall goal of optimizing the process as a whole.

3 CPS for Mobile Machines

After having described the motivation for CPS in mobile machines, this section presents an example of the general structure of a CPS in agricultural and construction machinery. The system boundary for this CPS is shown in Fig. 2. It consists of physical and cyber elements that communicate through a connecting interface. In addition, the two external elements *Worker* and *Global Environment* have to be taken into account when working with CPS, as they constantly influence the system and are also influenced by the system's physical actions.

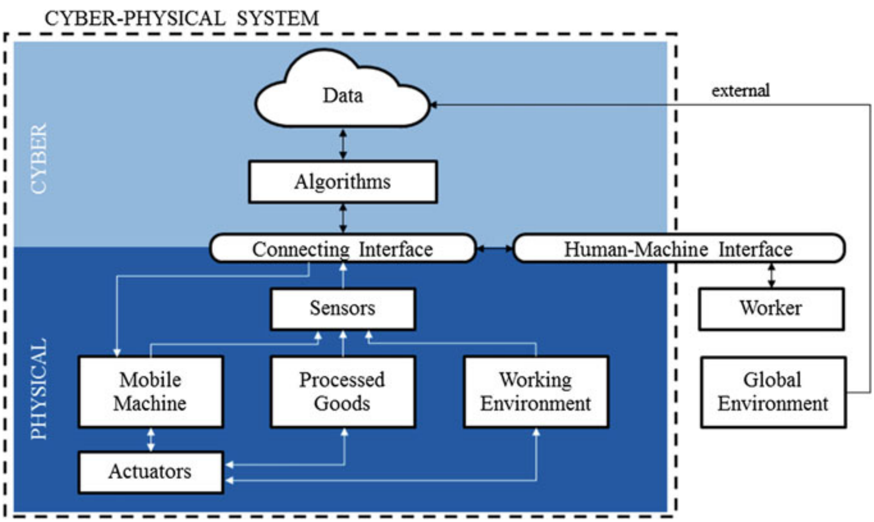


Fig. 2 System boundary of a CPS in mobile machines

The two cyber elements are the data and the algorithms. The *data* represents any kind of stored information. This storage may occur on a central server of the CPS or on a decentralized server network where not all elements need to belong to the CPS itself. The *algorithms* represent any processing of information made available through a bi-directional transfer from existing data or through a *connecting interface* (including the *Human-Machine Interface* (HMI)) from *sensors*. The physical part of a CPS in mobile machines consists of five elements. The core element is the *mobile machine* itself, which carries the *actuators*. The actuators adjust the mobile machine and enable it to interact with the *processed goods* as well as with the *working environment*. The mobile machine, the processed goods and the working environment are monitored by sensors which then use the connecting interface to deliver the obtained information to the cyber elements for processing. This connecting interface also allows the cyber elements to send instructions to the actuators of the mobile machine.

In this article, the *global environment* as well as the *worker* are considered external elements. The global environment is not detected by sensors of the mobile machine, because these are limited to the working environment. However, larger and potentially even global entities such as weather satellites or road traffic monitoring systems deliver data which is then available within the CPS and can therefore be processed by the various algorithms of the CPS. The worker can only communicate with the mobile machine through a HMI [18] which is linked to the connecting interface of the CPS as defined in Fig. 2. Which of the physical elements are seen and processed by the worker, highly depends on each specific system, which is why aside from the HMI no connecting arrows lead to or from the worker in the presented definition of boundaries. Most of the working environment, such as an obstacle ahead or the arrival of transportation vehicles can be seen by the worker at all times. Other information such as the vehicle speed can only be assessed but not known exactly by the worker. Still other information such as the current fuel consumption cannot be perceived directly at all. This information is only accessible through monitors, which show the information obtained from sensors. These monitors are therefore an example of a HMI.

4 Data

In Sect. 3, *data* was defined as a cyber element of a CPS and represents any kind of stored information. This storage may occur on a central server of the CPS or on a decentralized server-network where not all elements need to belong to the CPS itself (Fig. 2). The stored data can be of permanent or transient nature [55], whereas the transition is gradual.

Some data can be considered constant, such as basic geometrical and technical information about the machine, geomaps containing information on the elevation, texture and slope of the field (agriculture) or 3D data of planned surfaces (construction). In contrast, machine data like vehicle and tool positions, velocities, motor speeds and current fuel consumptions belong to transient data. Furthermore, crop and soil properties such as moisture content, diseases of crops and nitrogen status can also be regarded as transient data in agricultural systems. There is a lot more information that is relevant in the field of mobile machinery, like weather conditions and fuel and market prices. Those data, however, cannot be classified as purely transient or purely permanent, but can be arranged somewhere between these two extremes.

The allowable time delay in the communication can be defined on the basis of this time-dependent classification. While permanent data is not critical to communication time, real-time communication may be important for transient data. However, a real-time communication of all transient data to the database would lead to an enormous amount of data transfer and is therefore not realistic. In fact, relevant data should only be communicated when necessary. This, however, is highly dependent on the data relevance, which is mainly determined by economical and security issues.

Furthermore, in order to define communication interfaces and protocols, the level of data acquisition and the level of data usage have to be considered. In Table 1, six different data acquisition and data usage levels are listed along with typical data examples in the field of agricultural and construction machinery.

These levels can be divided into global, regional, business, site, machine and actuator levels. While the former levels include very general information relevant for the overall processes, the latter levels include very detailed data regarding specific applications and components. As a result, the complexity of communication very much depends on the difference between these levels. As an example, the

Table 1 Different levels of data acquisition and usage

Level	Construction	Agriculture	Data examples
Global	Internet	Internet	Market prices of processed goods
Regional	Country, state	Country, state	Law, weather, fuel prices
Business	Intranet, company	Intranet, farm	Fleet logistics, customer orders
Site	Construction site	Field, barn, plantation	Geomaps, soil properties, weather
Machine	Construction machine	Agricultural machine	Fuel, wear, workload, yield
Actuator	Drivetrain, machine equipment	Drivetrain, machine equipment, winch	Load, temperature, wear

usually applied CAN bus for communications within the machine cannot be used for transferring data from the machine to the entire construction site. Hence, additional interfaces and protocols have to be used, increasing the complexity of communication significantly.

A change in communication is therefore highly probable to handle the rising complexity due to the continuously increasing amount of sensor data on different levels (see Fig. 3). At the moment, most of the communication is based on a classical, hierarchical system with a central element. A typical example is the machine control system, which—among others—collects actuator data, processes it and finally sends signals back to the actuator. This way of communication, however, is not very efficient and more complex communication strategies including small and intelligent, interconnected subsystems can be used. A typical example is the combination of existing sensor data to generate more information about the system than the sum of all sensor data could deliver when evaluated individually. In the future, these intelligent interconnected subsystems will increase and result in a fully connected and decentralized system without hierarchical orders. This will enable the efficient usage of complex data and avoid redundant and inefficient data transfer.

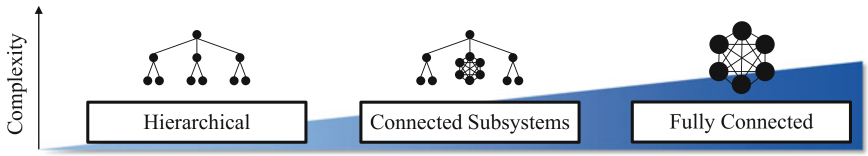


Fig. 3 Communication strategies for different complexity levels

5 Key Technologies

As mentioned in Sect. 2, Cyber-Physical Systems are the enabling technology for the fourth industrial revolution. However, CPS cannot be considered a singular invention. In fact, they are based on the existence of a few technologies which need to be smartly combined to enable the creation of Cyber-Physical Systems. These technologies are therefore considered key technologies and will be described in this section. Each key technology will shortly be discussed in the context of agricultural and/or construction machinery. Some key technologies also involve data processing and could therefore be considered key algorithms (Sect. 7). However, throughout this publication, key algorithms are defined as algorithms benefiting specific issues in the mobile machinery sector. Hence, data processing within a technology is considered part of the key technology itself.

The first key technology that enables CPS in mobile machinery is *sensor technology* [19, 28, 98]. Various sensor types can deliver a wide range of information. The most common class of sensors delivers information on basic machine states such as the position of actuators, hydraulic pressures and flow rates, forces and motor temperatures [95]. The second category of sensors enables machine vision technology, such as sensors on the vehicle that gather data about the environment. This category includes radar [62] and ultrasonic sensors [71], measuring distances by calculating the time of flight of the signal wave between the transmitter and the reflector and back. Optical sensor systems can be found here as well [11], also evaluating the time of flight or processing the optical images directly. Machine vision through sensors could also allow new machine concepts, as the driver does not need to be able to see the entire machine environment directly. Worksite safety regulations only demand that the driver has a sufficient overview at all times. Camera or sensor-based systems are explicitly permitted [34, 60]. In agriculture, the machine vision sensor category is especially important in viticulture [98]. A third category consists of remote sensors. These sensors use electromagnetic radiation, infrared reflection or electrical resistance [55]. Spectrometers and fluorescence (remote) sensors are applied especially for weed mapping [43]. A fourth category of sensors is formed by independent robots [8] such as unmanned aerial vehicles (UAVs) [46, 69], known as drones, but also ground-based vehicles carrying sensors, such as robots for scouting and monitoring [3, 49, 73]. With vehicles becoming larger and more dynamic, safety is an increasingly important issue. Part of this issue can be tackled by machine vision technology outside the vehicle, but attention must also be paid to the driver through driver fatigue and distraction recognition systems [88]. To make them applicable in mobile machinery, one very important attribute of sensors is robustness, as they permanently need to withstand harsh environmental conditions.

The second key technology consists of *positioning systems* [28, 98]. They allow the localization of a machine or vehicle within a defined environment, for example globally, on a construction site or on a field. Most positioning systems are based on distance measurements. The corresponding physical sensor technologies were named in the previous paragraph.

The key technology of positioning systems consists of the well-established evaluation of distance signals and the conversion of data into position information. The applied physical principle is generally based on the measured distance and therefore on the carrier types of transmitter and receiver or reflector. Positioning systems in this paragraph are therefore grouped by these carrier types. One example of positioning systems are global navigation satellite systems, short GNSS [71] with satellites constantly transmitting signals and therefore being the carrier. Various methods exist to improve the accuracy of those systems, for example Real-Time-Kinematics (RTK), but also methods involving additional terrestrial reference stations [50]. Terrestrial systems can also be used solely to determine a position [13]. However, the manual effort with terrestrial systems is higher as the carrier stations need to be set up and in some cases moved regularly. Another method to obtain relative position information is the optical recognition of surrounding objects, e.g. buildings or landmarks via photogrammetry [89].

The third key technology are *auto-guidance systems* [98]. In the context of agricultural and construction machinery, two types of auto-guidance systems are relevant, enabling the automatic and precise execution of a digitally calculated and optimized movement. Most common are auto-steering systems, controlling vehicle movement [28, 61, 82] based on a preprogrammed or automatically calculated route and information of global and local positioning systems. However, as agricultural and construction machines are usually equipped with multiple actuators, such as shields, shovels or agricultural attachments, the control of these actuators might even be more important, especially if many actuators need to be controlled simultaneously. Automatic actuator control systems have therefore been implemented in many applications [7, 38, 67, 71, 101].

The fourth key technology is *mapping*. The general goal of mapping is to obtain information on a specific area with the necessary resolution and content. This information is then used in various key algorithms (Sect. 7) to optimize a process. The information obtained by mapping needs to be of adequate quality with respect to the specific process. In mapping systems, there are different demands for the agricultural and construction sectors. In the agricultural sector, the processed goods —for example soil or fertilizer—are usually more relevant than in the construction sector. Mapping in agriculture is therefore usually referred to as *geomapping* [28, 55] and aims at gathering information on the properties of these processed goods within a designated area rather than information on the area itself. In construction machinery, the most important information is usually the height of the surface rather than information on the processed material, e.g. soil or gravel. As mapping and geomapping combine position information with discrete values, such as soil properties or surface height, they also combine the corresponding technologies of the positioning systems and sensor technology. Like positioning systems, mapping solutions can be grouped according to the sensor carrier which, can be ground-based, vehicle-based, airborne and spaceborne. In general, these systems are suitable for different applications with respect to the area size to be mapped and the required accuracy of the mapping data, where the accuracy declines with increasing area size [89]. Years ago, mapping was carried out with terrestrial stations by adding all manually measured points into a map [13] and calculating surface slopes. Today, using a ground-based vehicle carrying one or multiple electronic positioning systems is a compromise between good area coverage and sufficient accuracy [53]. If areas above 100,000 m^2 need to be mapped, airborne or even spaceborne systems are applied [89]. In recent years a new technology has thus entered the market. Unmanned aerial vehicles, short UAVs and commonly referred to as drones, offer surveillance and/or measurements of larger areas while offering accuracy in the range of about two centimeters [75].

The fifth, and very important, key technology to enable Cyber-Physical Systems in mobile machinery is the *integrated information and communication technology*, commonly referred to as integrated ICT [28], enabling data and information transfer between all entities within a CPS [14, 66, 86, 112]. Various solutions exist in the market for this transfer of information, ranging from WLAN and Bluetooth connections transferring data to a local control station up to satellite communication [74] transferring data to a cloud- or server-based system [90]. To enable the communication of sensors, components, controllers and servers of different producers, standardized communication protocols have been under research in the past years [78]. HMI also play an important role in this technology, as they form the only way in which digital information can be made accessible to an operator and vice versa.

Another key technology, albeit only important in the agricultural sector, is *variable rate technology* (VRT), allowing precise seeding and optimized application rate efficiency for spraying (nutrients, herbicides and pesticides) according to the variation in soil nutrients or plant growth [28]. An algorithm evaluating geo-maps and the current machine position on the field determines the applied rates. VRT reduces costs as well as the environmental impact [98].

6 Key Algorithms

In the last decades, the amount of digitally produced data has significantly increased and in 2014, machines generated more data than humans for the first time [79]. Today it only takes a single day to generate the total amount of data produced by the year 2000 [83]. However, simply creating and collecting data does not create benefits [25]. Data must be transferred to the right place and then be processed and combined by algorithms in a smart and efficient way. Algorithms are thus the core of all strategies and processes in agricultural and construction machinery. The following section describes the most important key algorithms for construction and agricultural machinery.

The first key algorithm is *job planning*. It lays the foundation for all work steps as it provides digital information on all process steps, machinery and worksite properties, among other information. By having up-to-date information on all ongoing processes, machinery positions and utilization data, it is possible to derive the current job progress. When this progress is compared with the initial schedule, delays can be detected and countermeasures initiated. This allows control over the progress. In general, job planning tools also offer job control functionalities [2, 45, 108, 110, 111].

The second key algorithm is *navigation*. Using positioning data as well as information obtained from job planning, navigation algorithms can be used to determine a desired driving path. The actual vehicle movement along this path is then executed with the help of the key technology auto-steering. Compared to navigation systems known from the automotive sector, the precision requirements in agricultural and construction machinery are much higher. An example of a navigation algorithm from the construction sector is lane planning during the asphalt compaction process [52]. Here, speed and steering are controlled simultaneously to optimize the compaction result. The use of modern navigation systems led to this key algorithm in agricultural machinery, where it is called *controlled traffic farming* (CTF) [44]. CTF is a management tool that "divides the crop area and traffic lanes into distinctly and permanently separated zones. All implements have a particular span … and all wheel tracks are confined to specific traffic lanes" [29]. Setting up evenly spaced permanent traffic lanes requires an accurate guidance system. Currently the best guidance system for high repeatability is the Real Time Kinematic (RTK) GNSS Autosteer with an accuracy up to ±2 cm. Benefits include less subsurface compaction, leaving 80–90 % of the field area without compaction [98], improved input efficiency from less overlap and less fuel use from running on firm tracks. CTF also enables further technologies such as shielded or banded spraying, inter-row sowing, easy on-farm trials and relay planting [59].

For *fleet and job management*, different machines involved in a process (e.g. harvesting) or job (e.g. excavation) are crosslinked and communicate with each other in order to improve the process. An algorithm ensures optimal travel distance, working time and fuel consumption. Modern fleet and job management algorithms lead to more automation. Further developments and involvement of greater CPS will enable autonomous, self-optimizing processes [91]. Fleet management in particular also involves vehicle disposition [6, 20, 85] as well as the scheduling and tracking of maintenance intervals [74, 108] to enable higher machine utilization. In this context, current research is in progress to use Cyber-Physical Systems to improve the scheduling of proactive maintenance [31].

Smart m*achine operating strategies* are also a key towards improving the working processes and therefore increasing productivity and efficiency. These strategies consider single machines (including small CPS). Current examples are the intelligent adaptation of the motor speed to the current demand or the mechanical disconnection of power consumers when they are not needed. Many more examples of smart operating strategies for the optimization of single machines could be given at this point. In the future, however, these strategies will increasingly consider the fleet and process as a whole [26] by having all necessary information available at all times through the key technology of integrated ICT. Evaluating the efficiency of smart machinery will also lead to smart evaluation methods in the future [27].

Due to the greater importance of the processed goods in the agricultural sector, some additional algorithms only exist for agricultural machinery. These are described in the following paragraphs.

Precise irrigation management enables water conservation and the protection of plants or crops. It requires accurate monitoring of the soil's water status "over time in representative locations in the field" ([35], pp 2) in order to provide the right amount of water at the right place at the right time. Examples for precise irrigation management are Subsurface Drip Irrigation (SDI), Regulated Deficit Irrigation (RDI), Center Pivot Irrigation and Partial Root Drying (PRD) [98]. Remote sensing through manned machines or UAVs acquire high-resolution (30 cm) thermal imagery, providing maps of the site-specific variation of water stress [98].

Site-specific soil cultivation is for primary and secondary cultivation as well as for seedbed preparation. The precise control of working depth and clod size reduction are essential for a profitable result. Algorithms combine data from geo-maps and signals to control the working depth in real time [55].

Site-specific sowing ensures the exact amount of seeds needed to be sowed on arable land for an optimal yield and a high sustainability. Algorithms use geomaps (e.g. yield maps of the previous season, maps of soil texture) combined with positioning data in order to control the seed density in real time. Advanced working attachments for sowing ensure a precise adaptation of the requested seed density using electric drives [55].

Site-specific spraying, including fertilizing, weed control and pest control, provides a substantial and economical spraying process, deploying the optimal amount of fertilizer, herbicide or pesticide. The algorithms include either information of previously generated geomaps combined with positioning systems or real-time information from sensors applied on the sprayer. The applied resolution is essential for economical site-specific spraying. Optimal cell sizes for the distribution of spray and the applied sensors have to be determined [55].

Site-specific weed control offers the potential to reduce herbicide. Today an offline-approach and a real-time approach are state of the art. The offline approach relies on previously generated maps of weed distribution, while the real-time approach combines patch spraying technologies and real-time data from sensors such as coverage, yield loss effects by weed species and weed density. Species-specific weed control can be expedient as well. In the future, robots might control weed by either hoeing or applying herbicides [43].

Real-time yield monitor and mapping shows and records data such as mass and volume flow or the moisture content of the grain. Harvesting machines are equipped with sensors, measuring for example the force, the impact of the harvested goods or the absorption of gamma rays [32]. The site-specific recording of the yield also delivers the basic information needed for the above-mentioned key algorithms in agriculture.

7 Exemplary Processes

The current state of the art of CPS in the construction and the agricultural sector will be presented with the help of two selected exemplary processes. Examples of the previously introduced key algorithms will be shown and the enabling key technologies put in context.

7.1 *Construction Process*

In this section, the earthworks of a highway segment construction will be described as an exemplary process. This process is divided into six consecutive and two parallel work steps. These steps are shown in Fig. 4. They can be further divided into preparative (1–3) and executive (4–6) work steps. In addition to the earthworks, some examples regarding the processing of additional layers in street construction are discussed.

The first step in the exemplary construction process is *3D-planning*.

None of the previously described key algorithms and key technologies are used in this work step, as it is solely based on the existence of a 3D planning software.

This step involves the planning of the surface geometry as well as the material layers and their thicknesses to represent the final product, a highway segment. Even today, many construction projects are still planned in 2D, making precise planning afterwards very complicated [51]. In this case, digital 3D planning is inevitable, establishing the basis for all following steps to be carried out [40]. The planning of

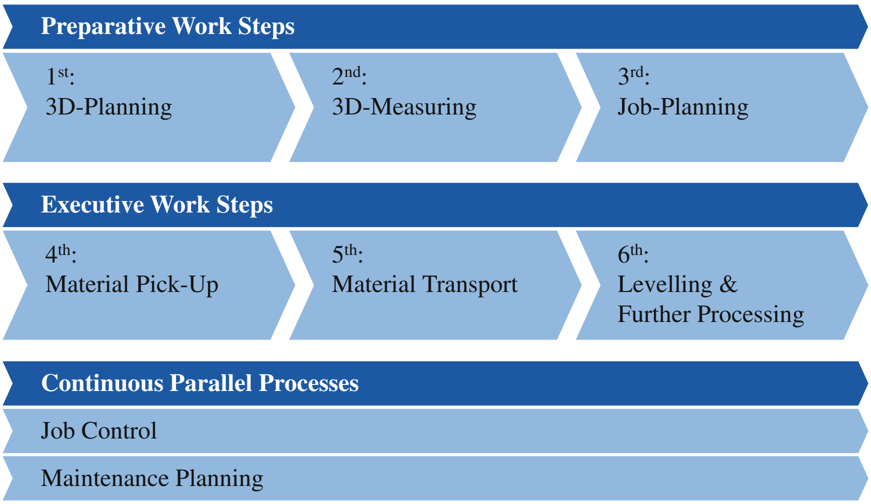


Fig. 4 Work steps during the earthwork of a highway segment construction process

the general course of a highway also depends on the shape of the existing surface in order to minimize the total volume of soil to be moved. However, at this point, only rough information on the existing surface is necessary.

The second step in the described process is *3D-measuring*.

Key algorithms used: *Navigation.*

Key technologies used: *Sensor technology*, *positioning systems*, *auto-guidance systems*, *mapping*, *integrated information and communication technology.*

After defining the exact geometry to be produced, the current status of the surface in the designated area needs to be determined. Various systems can be utilized for this step. Possible sensor technologies—such as distance measuring via laser or radar sensors, processing of optical images or the enhanced evaluation of GNSS signals as well as different sensor carriers—including hand-held devices, ground-based vehicles, UAVs and satellites—were introduced in Sect. 6. With progress in sensor technology, UAVs are becoming increasingly important in the construction sector [63, 97], offering accuracies of around two centimeters [75]. More and more companies are entering the market, offering to map an area of interest as a service [53]. Manufacturers of construction equipment have recognized this development and thus have already started to partner with manufacturers of UAVs and providers of systems that process the corresponding data to offer customized mapping solutions for the construction industry [45]. The described step of mapping an area of interest might be carried out repeatedly during other work steps, e.g. material pick-up and surface leveling, perhaps even continuously to obtain information on the current 3D surface. Only through a constant survey of the ongoing work in those steps can the full potential of precise planning be tapped [40].

The third step is *job planning*.

Key algorithms used: *Job planning*, *fleet and job management.*

Key technologies used: *integrated information and communication technology.*

The previous work steps served to determine the current and planned surface geometry as well as the thickness of the layers to be added. With this information, the job can be planned. For the basic earthworks, the height differences between the current and planned surface are calculated. From these it is possible to derive the volumes which need to be removed or added at each point. Years ago, this was done by manually calculating volumes based on the manually measured surface points. Today, modern software tools offer automatic calculations of these volumes [51]. Some tools even derive machinery utilization planning from these volumes, generating work plans and schedules for available machinery [66] and minimizing the total cost of the earthmoving process [40].

The next step in a highway construction process is *material pick-up*.

Key algorithms used: *Job planning*, *navigation*, *fleet and job management*, *machine operating strategies.*

Key technologies used: *Sensor technology*, *positioning systems*, *auto-guidance systems*, *mapping*, *integrated information and communication technology.*

After the theoretical and digital planning have been completed, physical earthmoving needs to be carried out. Material therefore needs to be picked up in order to be loaded onto a transport vehicle afterwards. For this task, various machine types exist today, for example excavators and wheel loaders. Due to the higher kinematic complexity of excavators, the automation of these machines has been under investigation in the past years, leading to multiple 2D [16, 65, 67, 101, 104, 109] and 3D [68, 80] control solutions and even integrated weighting functions [92]. Excavators usually do not travel regularly in their normal work cycle, which makes automatic machine control easier to implement. Through integrated ICT, alterations of the planned surface geometry can be transferred to the mobile machine quickly with low effort, allowing new plans to be executed quickly as well. Wheel loaders travel a lot more during their normal loading cycle. However, manufacturers of wheel loaders also offer partial automation systems. One product, for example, is the automatic filling of the bucket [21, 93] to allow full driver concentration on vehicle speed while observing the surroundings to improve worksite safety. These semi-automatic systems can also lead to reduced tire wear and increased machine productivity and efficiency [21, 93].

After this, *material transport* takes place as the next step.

Key algorithms used: *Job planning*, *navigation*, *fleet and job management*, *machine operating strategies.*

Key technologies used: *Sensor technology*, *positioning systems*, *auto-guidance systems*, *mapping*, *integrated information and communication technology.*

The material needs to be transported within the construction site or completely removed from it. For very short distances within the worksite, a wheel loader might be suitable. In general however, a separate transportation vehicle is necessary. To transport the material efficiently without interruptions and delays, a well-planned vehicle disposition is essential since, for example, an excavator or wheel loader cannot continue work without a transport vehicle in place and vice versa. To plan this disposition, a wide variety of software tools is available [6, 40, 84]. However, all these tools depend on having current information on vehicle positions, vehicle loading statuses, ongoing work, future capacity requirements and, if any transportation takes place on public streets, traffic information. Some of this information can be delivered from sensors integrated in the vehicles, usually the current position or the loading status [23, 92, 107] while traffic information, for example, is delivered by external traffic information systems. To keep this information up to date, machines communicate with each other [90] and GNSS position data is continuously transferred from each vehicle to the controlling system [6, 64, 70, 96]. Additionally, information on vehicle capacities, material properties and other boundary conditions need to be known by the system. If the transported material has time-dependent properties—such as the temperature of asphalt or the hardening status of concrete—timeliness in the logistic transportation chain is even more important. In this case, complex models need to be used, taking into account further information like ambient temperature or rainfall [6].

The final physical work step is the surface *levelling and further processing* of material.

Key algorithms used: *Job planning*, *navigation*, *fleet and job management*, *machine operating strategies.*

Key technologies used: *Sensor technology*, *positioning systems*, *auto-guidance systems*, *mapping*, *integrated information and communication technology.*

After excess material has been removed or missing material has been delivered, a defined surface height needs to be created. This may be a plane, a defined slope or a different specific shape. Planes are usually created with graders or dozers while excavators are used for other geometries. Today, multiple producers offer solutions customized for dozers and graders [22, 101] in addition to the above-mentioned 2D and 3D control systems. These systems need to have information on the geometry that has to be created as well as the exact position of the machine's actuators. While a leveling system on a dozer or grader only needs to know the position of the shield relative to the desired surface height, multiple sensors must be installed on a machine like an excavator with more complex kinematics, delivering information on the position of every link of the actuator [99]. Including the position of the entire vehicle, the absolute position of the shovel or shield can be calculated and from that the surface leveled or shaped surface can be derived.

In highway construction, but also in other processes, various material layers such as gravel and asphalt need to be added after the basic soil has been leveled. Furthermore, the thicknesses of these layers have to be produced as planned to meet national regulations [39]. The processing of these layers requires compaction in the previously completed leveling of each layer. This compaction needs to be carried out very precisely in order to produce a long lasting asphalt layer [86]. The asphalt can only be compacted while it still has a certain temperature [52], so the process is also time-critical. To optimize the result, machine control systems have been developed that offer precise lateral steering as well as speed control to ensure equal compaction in the area of interest [52, 71, 102]. As the width of the road paver is usually greater than that of the compaction machines, precise lane planning is also necessary and offered by some machine control systems [52].

The first of the two *continuous parallel processes* is *job control.*

Key algorithms used: *job planning*, *fleet and job management.*

Key technologies used: *Integrated information and communication technology.*

Job control describes the constant, ideally automatic, surveying of the ongoing work. By comparing the current status—such as the size of a leveled area or the excavated volume—to the planned value of the same parameter, progress can be monitored, delays identified early and any needed countermeasures can be initiated. Job control is a functionality that is usually integrated in the same tools that offer the initial job planning [17]. In the future, the initialization of countermeasures or at least the proposal of possible countermeasures will also be carried out by job control tools, allowing even shorter reaction times for unplanned events [91].

The second continuous parallel process is *maintenance planning*.

Key algorithms used: *Fleet and job management.*

Key technologies used: *Sensor technology*, *integrated information and communication technology.*

The benefits of maintenance planning apply to small and large worksites. On a small worksite, the failure of a single machine can lead to stagnation of the entire construction process as other work steps may depend on the output of this particular machine. On larger worksites, where many machines of the same type may be in use at the same time, precise maintenance planning can lead to significantly greater utilization of existing machinery and maintenance resources [81]. In both cases, it reduces the downtime of machinery, especially unplanned downtime. This is made possible through wear detection as well as wear and fatigue prediction. Many producers of construction machinery offer maintenance planning as part of their fleet management software tools [70, 94], as well as the possibility to quickly order spare parts straight from these software tools [24]. Also, companies outside of the mining and construction industry have recognized the potential of maintenance planning and started the development of the corresponding software tools [10].

7.1.1 Conclusion CPS in Construction Machinery and Future Potential

Each of the introduced work steps can be optimized using the key technologies and key algorithms. Therefore, commercially available systems and products have been described in each work steps. Most of these products aim at optimizing one particular work step while in recent years some research activities have also aimed at optimizing larger parts of processes. From the described process and the ongoing research in this field it can clearly be concluded, that further crosslinking, smarter data evaluation as well as the introduction of novel sensors will enable much higher process efficiency in the future by greater vehicle utilization, less machine downtime and much less manual planning and control effort. Also, automated or autonomous machines will deliver desired results quicker by minimizing the movement and processing of unnecessary volumes of soil, gravel and other materials.

The introduction and implementation of new hardware and software technologies in the construction sector will tap currently unused potentials and lead to the formation of increasingly complex and powerful Cyber-Physical Systems. Large investments are currently made in the key technologies mentioned above, for example in drone technology [41]. A recent study even suggests the use of glasses with augmented reality on construction sites [106]. The fact that a greater part of the construction industry believes in “big data” [30] and digitalization in general [58] as one of the main factors shaping the future of construction and that in a recent study 86 % believe these systems will lead to much greater timeliness on construction sites [106] indicate how much future potential is offered when all these technologies are combined.

Machine manufacturers and suppliers have given some estimations on possible cost reductions through single measures. Connecting machinery alone can reduce fuel cost by more than 40 % [103] and move times of earthmoving equipment can be reduced by as much as 50 % through optimized planning [100], a key algorithm of CPS. On process-level, the Central Federation of the German Construction industry as well as the Komatsu corporation state it will be possible to finish construction sites 20–30 % more cost-efficiently [33, 77].

Between the years 2000 and 2011, the productivity of the German construction industry, for example, was improved by 4.1 %. However, the productivity improvement throughout the entire German industry, where many companies have started to engage in industry 4.0, was 11 % [5]. Applying the same digital transformation in the construction sector, might lead to a reduction of this difference in the future.

7.2 Agricultural Process

This section gives an overview of CPS in the agricultural sector. As introduced in the challenges section (Sect. 2.2), PA is a whole-farm CPS-based management approach to increase yields while ensuring ecological sustainability. According to [98], PA is divided into three applications:

- Precision Farming (PF) on arable land,
- PF within the vegetables, fruits and viticultures and
- Precision Livestock Farming (PLF).

This contribution focuses on PF on arable land since it "is the most widely used and most advanced amongst farmers" ([98], pp 15). Based on the crop growth cycle of CEMA [28] in Fig. 5, the key technologies will be put in context and further examples of individual technologies as well as whole CPS approaches in agriculture will be introduced.

The first step of the crop growth cycle is *soil preparation*.

Key algorithms used: *Controlled traffic farming*, *site-specific soil cultivation.*

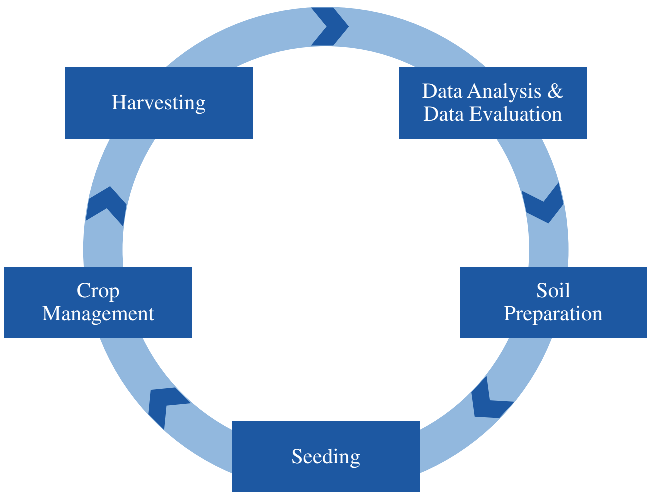


Fig. 5 Crop growth cycle (cf. [28])

Key technologies used: *Positioning systems*, *auto-guidance systems*, *integrated information and communication technology, geomapping.*

The main objective "is to produce a firm and weed-free seedbed for rapid germination and emergence of the crop" ([28], pp 7) by tilling or ploughing. Two cultivation steps are performed during soil preparation. The objective of the first cultivation step is to control the working depth. An algorithm therefore controls this working depth based on data of clay content, organic matter content, hydromorphic properties and the slope of the soil as well as penetration resistance measured online. The objective of the second cultivation step is clod size reduction. For this, an algorithm uses forces measured online on working tools as control parameters. An example of emerging single key technologies is the combination of CTF and site-specific cultivation, supplementing a CPS enabling the use of unmanned farm machinery or robots in future.

Soil preparation is one of the most energy-consuming steps in the crop growth cycle. PF helps to decrease fuel consumption and save time. An example of this is strip-tillage. "Strip-tillage creates narrow-width tilled strips … to increase … soil evaporation and soil temperature in the top two inches" ([1], pp 1) instead of ploughing the whole surface. As a result of strip-tilling, only 20 % of the soil is cultivated compared to ploughing. This requires less operations, saving time and fuel [28, 37]. Further advantages are reduced self-propagation between rows and erosion due to organic material retained on the surface and retained humidity as well as rain infiltration between rows [28].

The second step of the crop growth cycle is *seeding*.

Key algorithms used: *Controlled traffic farming*, *site-specific sowing.*

Key technologies used: *Variable rate technology*, *geomapping*, *positioning systems*, *auto-guidance systems*, *integrated information and communication technology*, *sensor technology.*

The main objective of PF in seeding is increasing yield and saving seeds. Seed rate or seed density are controlled by an algorithm based on geomaps (e.g. soil texture). Using the VRT, the determined seed density can be applied on the field. The sowing depths can be controlled either based on soil texture or on water content, depending on regional and climate characteristics. For soil texture control, the algorithm uses geomaps and distances measured online (e.g. by ultrasonic sensors). For water content control, the algorithm uses data of the soil moisture, measured by using infrared reflectance or electrical resistance. Actuators adapt the coulter pressure, which is measured online. This technique is not technically mature yet. Therefore, a recent research project "investigate[s] the relationship of seeding depth and applied downforce to soil properties in order to specify required control system response to field conditions" ([42], pp 139). Combined with positioning systems, data measured online can be processed as geomaps.

Site-specific sowing offers a uniform growth rate of the crops and minimizes the use of fertilizer during the third step of the crop growth cycle.

The next step is *crop management*. During this step, the crop needs irrigation, fertilization and protection. CPSs make it possible to carry out these aspects site-specifically and enable a higher yield with less input.

Key algorithms used: *Controlled traffic farming*, *site-specific spraying*, *site-specific weed control*, *precise irrigation management.*

Key technologies used: *Variable rate technology*, *geomapping*, *positioning systems*, *auto-guidance systems*, *integrated information and communication technology*, *sensor technology.*

Sensing of crop properties is inevitable in order to manage a crop site-specifically. There are different sensor technologies, using visible and infrared reflectance (to estimate the chlorophyll concentration within leaves), for example, or infrared reflectance and thermal radiation (to gain site-specific information on the water supply of crops).Yield prediction is possible and crop management can be improved if the crop properties are recorded repeatedly during the growth season. With proximal and remote sensing, it is possible to record data of larger areas and in shorter time intervals. By applying proximal sensors to mobile machines or robots, site-specific control is possible in real-time. An example is a sprayer equipped with optical sensors for the spot-spraying of weeds [55]. Examples of robots equipped with proximal sensors are a robot for scouting and monitoring used by Hohenheim University [49] and the BoniRob of Amazone for plant phenotyping [3, 73]. Remote sensing is possible from satellites or unmanned aerial vehicles (UAVs). Sensor-equipped, affordable micro-UAVs in particular are advancing in agriculture [46] and are already available on the market, such as the "PrecisionHawk" or the "eBee" [87].

Site-specific spraying—including fertilizing (spraying of nutrients) [55], weed (spraying of herbicide) [43] and pest control (spraying of pesticides) [98]—has a major impact on sustainable and economical farming. A "uniform application of fertilizer may not provide the right amount of nutrients to certain areas of the field, carrying the risk of over- or under-fertilizing plants" ([28], pp 12). For example, an infrared light sensor applied on a sprayer measures the amount of nitrogen in the leaves. An algorithm including additional data, such as yield-mapping data from the previous harvest, and an actuator ensure the optimal amount of nutrients on the go [55]. When this is combined with additional CPS key technologies, crop management can be made more precise. For example, the overlap of sprayed areas on a field can be decreased through automatic steering and by using GPS data.

Precise farming technologies also enable site-specific irrigation, such as subsurface drip irrigation (SDI) [28] or center pivot irrigation systems [35]. "Effective irrigation management for irrigation systems requires that soil water status be accurately monitored over time in representative locations in the field" ([35], pp 2). New technologies such as wireless underground sensor networks for soil moisture measurement do not need to be removed before harvest and therefore enable a more autonomous process.

The last step of the crop growth cycle is *harvesting*.

Key algorithms used: *Job planning*, *navigation*, *controlled traffic farming*, *fleet and job management*, *real-time yield monitor and mapping.*

Key technologies used: *Positioning systems*, *auto-guidance systems*, *integrated information and communication technology*, *mapping*, *sensor technology.*

This step can be the most critical one for the farmer. Since timing, speed and accuracy are essential here, CPS form the key for an improved harvesting process and a high yield.

Different machines for transport and harvesting have to be managed. CPS offer the crosslinking of these machines with sensors, integrated communication tools and geopositioning systems. Fleet management algorithms evaluate the information from all vehicles and process states in order to reduce travel distances, fuel consumption and working time. In addition, an extensive documentation of the working process is provided.

Another vision is the use of small unmanned machines. A concept for small individual combine harvesters operating as a remote-controlled swarm was investigated in [57]. Those new machine concepts enable swarm operations, providing further advantages such as robustness (due to redundancy of swarm elements) and safety (less powered vehicles cause less damage in case of control failure). These concepts also lead to less soil compaction.

Another important aspect is the recording of yield. Modern combine harvesters offer yield monitoring, showing the grain mass flow and the moisture content of the grain. Yield maps can be created when this is combined with positioning systems. Those can be used for crop rotation strategies and to determine site-specific seed and fertilizer rates [55]. Real-time monitoring technologies also enable a real-time operator performance analysis in order to identify potentials for performance optimization. As a result, the driver can be provided with suggested corrections for the machine operation right away [76].

7.2.1 Conclusion CPS in Agricultural Machinery and Future Potential

Each of the technologies mentioned above is state of the art and forms a CPS itself. They can also be used in combination to improve farming by forming even greater—though more complex—CPS. PF depends on information-intensive technologies. Technologies will therefore become more information-intensive in the future and will also need to be handled in a smarter way. "The main problem with … [PF] is selecting suitable sensing principles and appropriate processing methods for its signals" ([55], pp 2). This is where PF becomes Smart Farming (SF). "SF is based on Precision Farming but also dependent on information-intensive technology. The difference from previous systems is that this approach takes all areas of information into consideration and also has a higher level of knowledge and automation" ([47, 48], pp 25).

As shown above, the key technologies of CPS are applied throughout the whole crop growth cycle and improve productivity, ensure sustainability and provide economic benefits. In the future, these single key technologies will be crosslinked and merged to a greater CPS. This will lead to more automation and, further on, to more autonomy. A future vision could be an autonomously processed crop growth cycle.

CPS in agriculture generates great future potential.

Taking a look at specific algorithms shows the benefits gained by applying CPS in agriculture. According to [28] site-specific soil cultivation reduces emissions up to 50 % and increases productivity by 5 %. Site-specific spraying enables fertilizer savings up to 14 % and increases the average yield up to 6 %, creating an economic benefit to the farmer of 50–110 €/ha [28]. A study of [98] proved an increase of 25 % for water productivity by applying precise irrigation management during summer.

Applying CPS in agriculture will lead to improved processes and business models altogether. The increase in productivity is estimated by [4] to 15 % by 2025.

At present, big agricultural companies and large farms particularly benefit from CPS and dominate the current market. With further establishment of CPS, smaller farms and companies will also benefit from these new technologies. A whole new industry sector might be established, arising from start-ups. Today numerous recently launched start-ups involved in the mapping industry [54] or UAV technology [15] are gaining ground. It is also expected that new companies, focusing on the combination and merging of the above-mentioned single technologies, will appear and strengthen their market share. In 2014 big data companies, such as Google, invested more than $ 2.4 billion in agricultural technology start-ups [36].

It is difficult to calculate the market potential of CPS today, although it can be estimated by considering the market growth of PF. Research achievements in PF have come to practice within the past ten years. Today more than 70 % of new farm equipment sold worldwide has some form of PF technology [28]. For the PF market, an industrial study predicts "a compound annual growth rate of 12.2 % between 2014 and 2020 to reach $ 4.55 billion" [72]. However, the global agricultural equipment market is growing slower by only 4 %. The biggest markets are to be found in the developed countries, with North America and Europe as key markets [36].

8 Conclusion

This chapter has presented an overview of the current applications and future potentials of Cyber-Physical Systems (CPS) in the agricultural and construction machinery sectors. While the main motivation for CPS, increased operating and process efficiency, is identical in both sectors, the driving forces differ. In agriculture, the continuously growing need for food production and the demand for sustainable cultivation require an adaptation to local ground and plant properties. In the construction sector, the complexity of construction sites is increasing and the crosslinking of the single value steps is becoming more important. Hence, the future construction process consists of highly dependent, simultaneously executed work steps, requiring quick adaptations to undesired circumstances. CPS are essential to cope with these new challenges in the development of mobile machinery. A typical structure describing Cyber-Physical Systems in the mobile machinery sector has thus been presented and the main elements have been described. A categorization of data in mobile machinery into time-dependency, data relevance and the level of data acquisition and usage has been performed and the necessary future transformation of communication strategies has been presented. Six key technologies enabling the creation of CPS have been stated and specified in the context of mobile machinery. These key technologies are *sensor technology*, *positioning systems*, *auto-guidance systems*, *mapping*, *integrated information and communication technology*, and additionally the *variable rate technology* in agriculture. Thereafter, ten key algorithms have been presented, which enable the definition of strategies by smartly and efficiently processing and combining data. These key algorithms are *job planning*, *navigation*, *fleet and job management* and *machine operating strategies*. Furthermore, *precise irrigation management*, *real-time yield monitoring and mapping* as well as *site-specific soil cultivation*, *sowing*, *spraying* and *weed control* are used in agriculture. At the end of this chapter, two typical processes—a street construction process and a crop growth cycle—are analyzed in detail and the applications of the key technologies and key algorithms are exemplified.

The analysis of these processes allows the outlook that inevitably, higher crosslinking of machinery as well as novel sensors will be introduced in both sectors in the future, leading to the formation of much larger and more complex Cyber-Physical Systems. This crosslinking and the resulting increased availability of data at multiple levels will allow the currently pursued smart evaluation of data to generate information of much higher value. Only through the creation of this high-value information and the possibilities of densely connected Cyber-Physical Systems, can process efficiency as well as process safety in both the construction and agriculture sectors be taken to a next level.

Based on estimations of construction machine manufacturers and suppliers, Cyber-Physical Systems may reduce the total costs for construction sites by 20–30 %. The presented algorithms are able to decrease the fuel costs by 40 % and reduce the move times of earthmoving equipment by 50 %. Comparing the productivity improvements of sectors already engaging industry 4.0, Cyber-Physical Systems may increase the productivity in the construction sector by a factor of three. For the agricultural sector, an increase in productivity of 15 % by 2025 is estimated. Furthermore, with the establishment of Cyber-Physical Systems, also smaller farms and companies will benefit from the presented technologies and a whole new industry sector might be established.

References

1. Al-Kaisi M, Hanna M (2008) Consider the strip-tillage alternative. In: Resources conservation practices. Iowa State University Extension, Ames
2. Autodesk (2015) Machine 2 machine: how smart apps monitor construction site and equipment for better project margins. http://www.seecontrol.com/machine-2-machine-how-smart-apps-monitor-construction-site-and-equipment-for-better-project-margins/. Accessed 3 Dec 2015

3. Bangert W, Kielhorn A, Rahe F, Albert A, Biber P, Grzonka S et al (2013) Field-robot-based agriculture: "RemoteFarming.1" and "BoniRob-Apps". In: VDI-Berichte Nr. 2193, LAND. TECHNIK AgEng 2013, components and systems for better solutions, agricultural engineering, Hanover, Germany, 8–9 November 2013. VDI Verlag GmbH, Düsseldorf, pp 439–446
4. Bauer W, Schlund S, Marrenbach D, Ganschar O (2014) Industrie 4.0—Volkswirtschaftliches Potenzial für Deutschland, Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. (BITKOM), Berlin, Germany
5. Baumanns, T, Freber, P-S, Schober, K-S, Kirchner F (2016) Bauwirtschaft im Wandel—Trends und Potenziale bis 2020, study by Roland Berger Strategy Consultants GmbH, Munich, Germany
6. BaustellenApp (2015) Softwaregestützte Baustellenlogistik BaustellenApp & Leanmanagement. http://www.baustellenapp.com/. Accessed 4 Dec 2015
7. Biniasch A (2015) Per Laser auf Niveau. In: Computer-automation, Issue 09/2015, Haar Germany
8. Blackmore S (2014) Farming with robots 2050. In: The Oxford farming conference, Oxford, UK, 6 January 2014
9. Blanchet M, Rinn T, von Thaden G, De Thieulloy G (2014) Think act—Industry 4.0—the new industrial revolution, how Europe will succeed, study by Roland Berger Strategy Consultants GmbH, Munich, Germany
10. Bloem J, von Doorn M, Dulvenstein S, Excoffier D, Maas R, van Ommeren E (2014) The fourth industrial revolution—things to tighten the link between IT and OT. In: VINT research report 3 of 4, Groningen, Netherlands
11. Blum F (2014) Sicher mit 3D. In: Ke Next im Fokus-Automatisierung, Issue II/2014, November 2014. Moderne Industrie, Landsberg, Germany, pp 3, 20–23
12. Bock T, Linner T (2015) Robot oriented design—design and management tools for the deployment of automation and robotics in construction. In: The Cambridge handbooks in construction robotics, 1st edn. Cambridge University Press, Cambridge, pp 7. ISBN: 978-1107076389
13. Bollmann (2015) Modelle und Systeme von Geodaten-Methoden der Erfassung topographischer Geobasisdaten. From lecture: Karthografische Informationsverarbeitung, Department of Cartography, University of Trier, Trier, Germany
14. Bomag (2014) Bomag telematic. In brochure: Bomag Telematic, Boppard, Germany
15. Boughton B (2014) Potential of unmanned aerial vehicles (UAVs) in the grains industry. In: A report for Nuffield Australia farming scholar, Moama, New South Wales, Australia
16. BPS (2013) BPS III machine control system—robotics for crawler and wheeled excavators, front shovel and backhoe. In brochure: Axiomatic Technologies, Lempäälä, Finland
17. Bundesministerium für Wirtschaft und Technologie (2013) Autonomik-Autonome und simulationsbasierte Systeme für den Mittelstand. In: Autonomik—Band 1, Berlin, Germany, pp 13
18. Bundesministerium für Wirtschaft und Technologie (2013) Mensch-Technik-Interaktion. In: Autonomik-Band 3, Berlin, Germany
19. Bundesministerium für Wirtschaft und Technologie (2013) Multimodale Sensorik-Konzepte der Umwelterkennung/-modellierung. In: Autonomik-Band 5, Berlin, Germany
20. Bundesministerium für Wirtschaft und Technologie (2014) AutoBauLog-Autonome Steuerung in der Baustellenlogistik. http://www.autobaulog.de/. Accessed 7 Dec 2015
21. Caterpillar (2007) 950H wheel loader. In brochure: Caterpillar 950 Wheel Loader, Peoria, USA
22. Caterpillar (2007) AccuGrade—grade control system for hydraulic excavators. In brochure: Caterpillar Grade Control System AccuGrade, Peoria, USA
23. Caterpillar (2008) Payload control system 3.0. In brochure: Caterpillar Payload Control System, Peoria, USA
24. Caterpillar (2012) It pays to know—CAT product link. In brochure: Caterpillar Product Link, Peoria, USA

25. Caterpillar (2015) Get more mileage from your telematics data. In: Equipment world—get the jump on your competitors, Peoria, USA
26. CECE and CEMA—Committee for European Construction Equipment and European Agricultural Machinery (2011) CECE and CEMA optimizing our industry 2 reduce emissions, Brussels, Belgium, 17 October 2011
27. CECE and CEMA—Committee for European Construction Equipment and European Agricultural Machinery (2015) Driverless tractors and intelligent excavators: how the digital revolution is changing farm and construction equipment as we know it. In: Smart regulation for smart machines, Brussels, Belgium, 23 September 2015
28. CEMA—European Agricultural Machinery (2013) Smart equipment for sustainable agriculture, precision farming: producing more with less, Brussels, Belgium
29. Chamen T (2007) Controlled-traffic farming as a complementary practice to no-tillage. In: John Baker C, Saxton K (eds) No-tillage seeding in conservation agriculture, 2 edn. Wallingford, Oxfordshire, UK, pp 236
30. ConstructionDIVE (2016) 3 Concepts that will shape the future of construction. http://www.constructiondive.com/news/3-concepts-that-will-shape-the-future-of-construction-1/417049/. Accessed 11 Apr 2016
31. CORDIS—Community Research and Development Information Service (2015) MANTIS—cyber physical system based proactive collaborative maintenance. http://cordis.europa.eu/project/rcn/198079_en.html. Accessed 12 Dec 2015
32. Demmel M (2013) Site-specific recording of yields. In: Heege H (ed) Precision in crop farming, site specific concepts and sensing methods: applications and results. Springer, Dordrecht, pp 313–329
33. DeWitt, A (2015) Komatsu, smart construction, creative destruction and Japan’s robot revolution. Asia-Pacific J Jpn Focus 13(5):2
34. DIN 474-DIN EN 474:2010 (2010) Erdbaumaschinen-Sicherheit-Teil 1: Allgemeine Anforderungen (DIN EN 474:2010)
35. Dong X, Mehmet C, Suat I (2013) Autonomous precision agriculture through integration of wireless underground sensor networks with center pivot irrigation systems. Ad Hoc Netw 11 (7):1975–1987
36. Dressler N, Gundermann S, Keese S, Aulbur W, Zhang J, et al (2015) Business opportunities in precision farming: will big data feed the world in future? Roland Berger Strategy Consultants GmbH, Munich
37. Dummett I, Deeks L, Rickson J, Moldon J (2015) Evaluating oilseed rape establishment under a range of strip tillage practices. In: VDI-Berichte Nr. 2251, LAND.TECHNIK AgEng 2015, Innovations in agricultural engineering for efficient farming, agricultural engineering, Hanover, Germany, 6–7 November 2015. VDI Verlag GmbH, Düsseldorf, pp 481–486
38. Elflein N (2014) Digitale Assistenten-Tempomat für Minenfahrzeuge. From: Siemens Picture of the Fututre-Das Magazin für Forschung und Innovationen, Munich, Germany, 1 April 2014
39. FGSV-Forschungsgesellschaft für Straßen- und Verkehrswesen, Arbeitsgruppe Infrastrukturmanagement (2012) Richlinien für die Standardisierung des Oberbaus von Verkehrsflächen-RStO12, Cologne, Germany. FGSV Verlag GmbH. ISBN 978-3-86446-021-0
40. ForBau-Bayrischer Forschungsverbund Virtuelle Baustelle (2010) Virtuelle Baustelle-Digitale Werkszeuge für die Bauplanung und-abwicklung. In: ForBau Abschlussbericht 2010, Munich. Erlangen-Nuremburg, Regensburg. ISBN: 978-3-941702-18-9
41. Forbes (2016) Drone startup airware raises $30 million, adds Cisco’s John chambers to board. http://www.forbes.com/sites/ryanmac/2016/03/31/drone-startup-airware-makes-larger-enterprise-push-adds-former-cisco-ceo-to-board/#278ca63daac4. Accessed 19 Apr 2016

42. Fulton J, Poncet A, McDonald T, Knappenberger T, Shaw J, Bridges R, Balkcom K (2015) Considerations for site-specific implementation of active downforce and seeding depth technologies on row-crop planters. Downforce by depth settings influence seeding depth. In: VDI-Berichte Nr. 2251, LAND.TECHNIK AgEng 2015, Innovations in agricultural engineering for efficient farming, agricultural engineering, Hanover, Germany, 6–7 November 2015. Düsseldorf, Germany. VDI Verlag GmbH, pp 139–145
43. Gerhards R (2013) Site-Specific Weed control. In: Heege H (ed) Precision in crop farming, Site specific concepts and sensing methods: applications and results. Springer, Dordrecht, pp 273–294
44. Grains Research and Development Corporation—GRDC (2013) Controlled traffic farming. In: Fact sheet, northern, southern and western regions, implementing a controlled traffic farming system, Canberra, Australia
45. Grayson W (2015) With drones and driverless dozers, Komatsu to begin leasing automated construction fleets. In: Equipment world. http://www.equipmentworld.com/with-drones-and-driverless-dozers-komatsu-to-begin-leasing-automated-construction-fleets/. Accessed 3 Dec 2015
46. Grenzdörffer G, Engel A, Teichert B (2008) The photogrammetric potential of low-cost UAVs in forestry and agriculture. Int Arch Photogram Remote Sens Spatial Inf Sci 31 (B3):1207–1214
47. Griepentrog H (2011) Smart farming: a new formula for success. Agrifuture 11:24–27. Autumn
48. Griepentrog H (2011) Smart Farming: Praxisreife Lösungen, Erfordernisse und Techniken für morgen. In: DLG-Pressemitteilung, Hanover, Germany, 14 November 2011
49. Griepentrog H, Jaeger C, Paraforos D (2013) Robots for field operations with comprehensive multilayer control. Künstliche Intelligenz 27(4):325–333. doi:10.1007/s13218-013-0266-z
50. Groves P (2007) Principles of GNSS, inertial, and multi-sensor integrated navigation systems. Ind Robot Int J 39(3). ISBN: 978-1-58053-255-6
51. Günther W, Borrmann A (2011) Digitale Baustelle—innovativer Planen, effizienter Ausführen. Springer, Berlin, pp 23–80. ISBN: 978-3-642-16485-9
52. Halbrügge C, Johanning B (2015) Automatisierte Asphaltverdichtung mit schemelgelenkten Tandemwalzen. ATZ Off-Highway 8(2):59
53. Hansa Luftbild (2015) 3D-mobile-mapping by Hansa Luftbild german air surveys, Münster, Germany. http://www.hansaluftbild.de/dienstleistungen/geoinformation-dienstleistungen/3d-mobile-mapping.html. Accessed 3 Dec 2015
54. Harris T (2015) Agribotix, agricultural intelligence, drone-enabled. In: ENVI analytics symposium, Boulder, Colorado, USA, 25 August 2015
55. Heege H (2013) Precision in crop farming, site specific concepts and sensing methods: applications and results. Springer, Dordrecht
56. Herlitzius T, Aumer W, Geißler M (2010) Potenziale der Hybridisierung in Landmaschinen. In: VDI-Seminar Landtechnik, VDI, Freising, Germany, 8 July 2010
57. Herlitzius T, Stelzer R, Kranke G (2012) New concepts for harvesters. ATZoffhighway 5 (1):68–77
58. IHK (2015) Study of the German chamber of industry and commerce: Wirtschaft 4.0: Große Chancen, viel zu tun. Das IHK-Unternehmensbarometer zur Digitalisierung. Deutscher Industrie- und Handelskammerverlag, Berlin. ISSN: 1863-883X
59. Isbister B, Blackwell P, Riethmuller G, Davies S, Whitlock A, Neale T (2013) Controlled traffic farming technical manual. Department of Agriculture and Food, Perth
60. ISO 5006-ISO 5006:2006 (2006) Erdbaumaschinen-Sichtfeld-Testverfahren und Anforderungskriterien (ISO 5006:2006)
61. Kanai O, Sarata S, Kurisu M (2006) Autonomous scooping of a rock pile by a wheel loader using disturbance observer. In: Proceedings of ISARC 2006, Tokio, Japan, pp 472–475
62. Kirsch F, Miesen R, Voissek M (2014) Precise local-positioning for autonomous situation awareness in the internet of things, microwave symposium (IMS) 2014, Tampa, Florida, USA. doi:10.1109/MWSYM.2014.6848674

63. Knight W (2015) New boss on construction site is a drone. MIT Technology review, 26 August 2015. http://www.technologyreview.com/news/540836/new-boss-on-construction-sites-is-a-drone/. Accessed 11 Nov 2015
64. Komatsu (2013) Komatsu wireless monitoring system—KOMTRAX. In brochure: Komatsu Komtrax. Komatsu Europe International NV, Vilvoorde
65. Komatsu (2014) Komatsu hydraulic excavator PC210LCi-10. In brochure: Komatsu Hydraulic Excavator PC210LCi-10. Komatsu Europe International NV, Belgium
66. Komatsu (2015) SMARTCONSTRUCTION—with a future. Form: Komatsu Report March 2015. Komatsu Europe International NV
67. Lehnen B (2015) Implementierung einer (Teil-) Automatisierung von Arbeitsspielen in die Steuerungssoftware von Hydraulikbaggern. In: 6. Fachtagung Baumaschinentechnik 2015, Dresden, Germany, 17–18 September 2015, pp 294–305
68. Leica (2015) Leica Geosystems- Baumaschinensteuerung. In brochure: Leica Geosystems, 15 June 2015
69. Leonida C (2015) MM's Top 10 technologies: 6—UAVs and drones. In: Mining Magazine, 30 April 2015. http://www.miningmagazine.com/equipment/exploration/mms-top-10-technologies-6/. Accessed 3 Dec 2015
70. Liebherr (2010) LiDAT—Das umfassende Fuhrpark-und Flottenmanagement von Liebherr. In brochure: Liebherr LiDAT. Liebherr-Werk Nenzing GmbH
71. Marx B (2012) An Extra Dimension. In: IVT international 2012—industrial vehicle technology, off-highway edition, Section 4 electronics, pp 145–147
72. Mathur A, Salama B, Barros E, Helle J (2015) Digital agriculture: improving profitability, accenture precision agriculture service and connected crop solution help the agriculture ecosystem fulfill its potential. Accenture PLC, Dublin
73. Michaels A, Haug S, Albert A, Grzonka S (2013) Vision-based manipulation for weed control with an autonomous field robot. In: VDI-Berichte Nr. 2193, LAND.TECHNIK AgEng 2013: Components and systems for better solutions, agricultural engineering, Hanover, Germany, 8–9 November 2013. VDI Verlag GmbH, Düsseldorf, pp 289–294
74. Modular Mining Systems (2015) Modular mining IntelliMine overview (video). http://www.modularmining.com/. Accessed 3 Dec 2015
75. Peterman V, Mesarič M (2012) Land survey from unmanned aerial vehicle. International archives of the photogrammetry, remote sensing and spatial information sciences, vol XXXIX-B1, XXII ISPRS Congress, Melbourne, Australia, pp 447–451
76. Pfeiffer D, Blank S (2015) Real-time operator performance analysis in agricultural equipment. understanding unused potential and ways to improve from day to day. In: VDI-Berichte Nr. 2251, LAND.TECHNIK AgEng 2015, Innovations in agricultural engineering for efficient farming, agricultural engineering, Hanover, Germany, 6–7 November 2015. VDI Verlag GmbH, Düsseldorf, pp 359–365
77. Pöllath, K, (2014) Bauwirtschaft: Forschung und Industrie—zusammen Zukunft gestalten, Zukunft Bau Kongress 2014, Bonn, Germany
78. Poppy, W, Unger, E (2000) CanOpen for mobile construction machines—an open communication network for control and automation. In: Proceedings of the 17th ISARC, Taipei, Taiwan, pp 1–4
79. Probst L, Frideres L, Pedersen B, Pedersen B, Caputi C (2015) Service innovation for smart industry—cyber-physical systems in the 'value network'. In: European commission case study 37, European Union, p 7
80. Prolec (2015) Digmaster Pro—2D/3D excavator machine guidance. In brochure: Prolec Digmaster Pro. Prolec Ltd, Dorset
81. Raymond James (2014) The internet of things—a study in hype, reality, disruption, and growth and associates. In: U.S. research industry report, St. Petersburg, Florida, pp 20–22
82. RGN (2015) Rio's remote-controlled mines come online. In: Resource global network, November/ December 2015, London, England, p 8
83. Russwurm S (2014) Die Zukunft ist digital. In: CKI Konferenz, Aachen, Germany, 11 December 2014

84. Sanladerer S (2004) Vernetzte Transportlogistik am Bau-Neue Wege zur Steuerung der Baustellenfahrzeuge. In: VDBUM Info, Issue 5/2004, Weyhe, Germany, pp 23. ISSN: 0940-3035
85. Sanladerer S (2008) EDV-gestützte Disposition mit Telematikeinsatz und mobiler Datenerfassung in der Baulogistik, Dissertation, Munich, Germany. ISBN: 978-3-9811819-2-0, pp 38–41
86. Schmid A (2015) Industrie 4.0: Staus auf deutschen Autobahnen sollen kürzer werden. SmartSite project, University of Hohenheim, 29 September 2015 (in press)
87. senseFly (2014) eBee—senseFly, the professional mapping drone. Cheseaux-Lausanne, Switzerland
88. Shemer N (2015) Caterpillar partners with seeing machines. In: Mining magazine, 15 September 2015. http://www.miningmagazine.com/technology-top-articles/caterpillar-partners-with-seeing-machines/. Accessed 12 Dec 2015
89. Siebert S, Teizer J (2014) Mobile 3D mapping for surveying earthwork projects using an unmanned aerial vehicle (UAV) system. Autom Constr 41:1–14 (Elsevier, Amsterdam)
90. Sontheim F (2015) Assistenzsystem für den Lade- und Transportprozess bei der Erdbewegung. In: 6. Fachtagung Baumaschinentechnik, Dresden, Germany, 17–18 September 2015, pp 340–350
91. Spath D (2013) Perspektiven der Dienstleistungs- und Arbeitsforschung, From: Proceedings of Dienstleistungsforschung und Dienstleistungspolitik, Düsseldorf, Germany, p 18
92. Spring (2015) Boomerang—dynamic weighing system for wheel loader and fork-lift. In brochure: Spring Machine Control Electronics Systems, Dormelletto, Italy
93. Steine+Erden (2009) Laden wie von Geisterhand-Der neue Cat 950H mit AutoDig. In: Steine+Erden, Issue 5, 2009, Langenhagen, Germany
94. Steine+Erden (2013) Hi-Mate: Webbasierte Maschinenüberwachung, Hyundai Heavy Industries Deutschland. In: Steine+Erden, Issue 5, 2013, Langenhagen, Germany
95. Steine+Erden (2015) Cat-Radlader der M-Serie mit integrierter Kontrollwaage. In: Steine +Erden, Issue 4, 2015, Langenhagen, Germany
96. Steine+Erden (2015) Cat-Flottenmanagement Product Link-Suchen war gestern, finden ist heute. In: Steine+Erden, Issue 2, 2015, Langenhagen, Germany
97. Stoller D (2015) Drohnen von Aerosense überwachen künftig Baustellen und Felder, 26 August 2015. VDI Verlag GmbH, Düsseldorf. http://www.ingenieur.de/Branchen/Luft-Raumfahrt/Drohnen-Aerosense-ueberwachen-kuenftig-Baustellen-Felder. Accessed 12 Dec 2015
98. Tejada P (2014) Precision agriculture: an opportunity for EU farmers—potential support with the CAP 2014–2020, in-depth analysis. Publications Office, Luxembourg City
99. Topcon (2011) X-22—wireless control systems for excavators. In brochure: Topcon Wireless control systems for excavators X-22. Topcon Corporation, Livermore
100. Trimble (2001) Trimble's positioning system for material handling applications dramatically increase productivity. Trimble New's Release, 12 December 2001, Sunnyvale, California, USA
101. Trimble (2014) GCS900 grade control systems. In brochure: Trimble GCS900, Westminster, USA
102. Trimble (2015) Trimble earthworks solutions. In brochure: Trimble Earthworks Solutions, Westminster, USA
103. Trimble (2016) Trimble connected site, Westminster, USA. www.connectedsite.com. Accessed 25 Apr 2016
104. Ubexi (2014) iDig—classic and touch systems. In brochure: iDig Touch, Bayswater North, Australia
105. United Nations (2014) The world population situation in 2014. In: A concise report, Oldenbourg, New York, USA
106. VDMA (2015) Industrie 4.0 im Industrieanlagenbau—Revolution oder Evolution. In: Study from the German machinery and plant manufacturing association, Frankfurt, Germany

107. Volvo (2014) Caretrack. In brochure: Volvo caretrack, Volvo Construction Equipment, Eskilstuna, Sweden
108. Volvo (2015) Care track. In: Volvo construction equipment. Volvo Construction Equipment, Eskilstuna. http://www.volvoce.com/constructionequipment/na/en-us/partsservice/caretrack_old/Pages/Introduction.aspx. Accessed 3 Dec 2015
109. Wagener M (2014) Aktuelle Entwicklung im Bereich der Baumaschinensteuerung. In: Fachbeitrag aus der Praxis 121, Offenbach am Main, Germany
110. Wenco (2015) Control your entire fleet of mining equipment with centimeter-level accuracy. In brochure: bench manager—machine guidance system, Richmond, Canada
111. Wenco (2015) A complete suite of tools to help your mine produce more but cost less. In brochure: wencomine—fleet management system, Richmond, Canada
112. Zeppelin (2015) Connected Worksite-Die vernetzte Baustelle der Zukunft. Zeppelin Baumaschinen GmbH, Munich. http://www.zeppelin-cat.de/produkte/connected-worksite.html. Accessed 3 Dec 2015

Application of CPS Within Wind Energy—Current Implementation and Future Potential

Paul Kunzemann, Georg Jacobs and Ralf Schelenz

1 Motivation

The working principle of wind turbines is easy to understand. Kinematic wind energy is converted to rotational energy. The generated rotation can either perform mechanical work or be converted to electrical energy. Changing application and constant development of wind energy can be seen in history. Some major milestones are shown in Fig. 1.

The potential of wind as a renewable energy source was discovered very early in Persia in the 7th century [1]. The wind mill was then imported later in Europe around the 11th century and used in agriculture as a water pump or flour mill. This was the beginning of the well-known Dutch wind mill. At the end of the 19th century there was a significant further development of wind energy. In 1887, James Blyth from Scotland developed the first electricity generating wind turbine for private application [2]. At the beginning of the 20th century many global wind energy projects arose. However, the level of performance remained relatively low—the electrical power output of wind turbines stayed below 100 kW [3]. The level of performance increased significantly during the oil crises in the early 70s, as renewable energy resources such as wind power gained importance.

P. Kunzemann (✉) · G. Jacobs (✉) · R. Schelenz (✉)
Center for Wind Power Drives (CWD), RWTH Aachen University,
Campus-Boulevard 61, 52074 Aachen, Germany
e-mail: paul.kunzemann@cwd.rwth-aachen.de
URL: https://www.cwd.rwth-aachen.de/home/

G. Jacobs
e-mail: georg.jacobs@cwd.rwth-aachen.de
URL: https://www.cwd.rwth-aachen.de/home/

R. Schelenz
e-mail: ralf.schelenz@cwd.rwth-aachen.de
URL: https://www.cwd.rwth-aachen.de/home/

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series in Wireless Technology, DOI 10.1007/978-3-319-42559-7_27

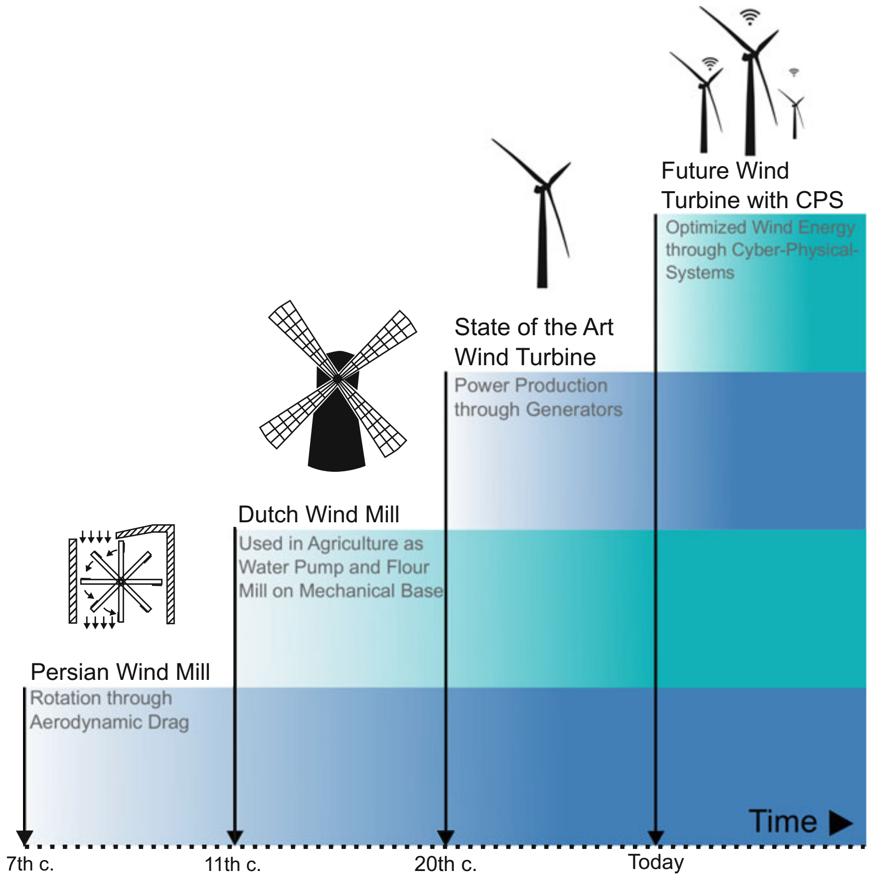


Fig. 1 Development of wind turbines

Nowadays, wind turbines operate in the MW power class with typically three blades faced in upwind direction. Benchmarks are 2.5 MW for onshore and more than 5 MW for offshore turbines [4]. It is common to operate the wind turbines in on- and offshore wind farms with plant sizes from several MW up to 630 MW, which is comparable to conventional coal and nuclear power plants [5]. This demonstrates the significance that wind energy has achieved in the current power grid. The rapid development and installation of wind energy can also be seen in Fig. 2, which shows the global cumulative installed wind power capacity from 2000 to 2014 [6]. On a global scale, 3 % of the electricity was already supplied through wind energy in 2013. In Denmark's energy balance of 2014, for instance, wind energy had a share of 40 % of the power supply [5]. According to the Global Wind Energy Council it is estimated, that 17–19 % of the global power demand could be supplied by wind energy by 2030 [6].

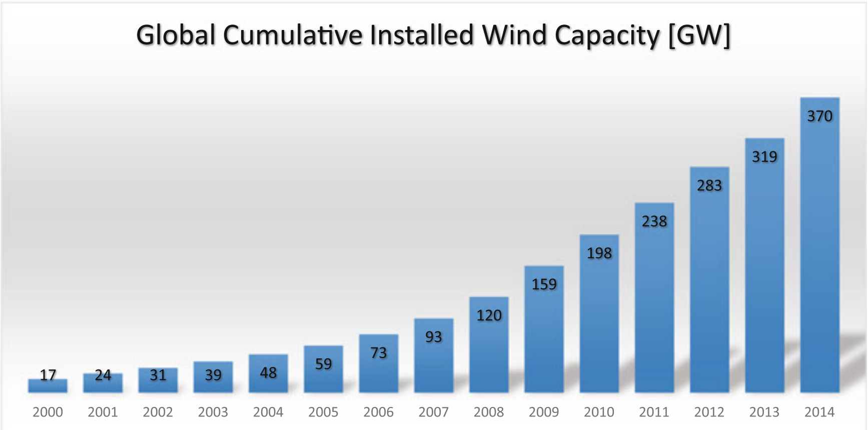


Fig. 2 Global cumulative installed wind capacity [6]

Political decisions have been one of the main drivers for this development in wind energy in the past years. In the course of the energy transition many regulations and targets were set to substitute the power supply from fossil fuels by renewables and mitigate the greenhouse gas emissions (GHG). The European Union, for instance, has the ambition for 2020 to increase the power supply by renewable energies up to 20 % and at the same time minimize the GHG emissions by 20 % [7]. Other countries have even more ambitious goals. For instance, Germany wants to reduce the GHG emissions by 40 % until 2020, compared to the emissions in 1990.

From all renewable energy sources, which are essential for the energy transition, wind energy has established as the key driver [8]. Today, wind energy is the world's fastest growing energy source and provides the potential to further substitute fossil fuels in the near future [4]. It is already competitive with conventional power plants. In Germany, for instance, the total cost of energy for onshore wind sites with good wind conditions is at least in the same range as coal and gas [9]. Due to this rapidly ongoing development and installation of wind energy, this technology provides high potential in the future. Even slight improvements and optimizations can have a major worldwide benefit, as wind energy contributes a significant amount of energy to the entire power supply.

To take advantage of the existing potential in wind energy, wind turbines operating as Cyber Physical Systems (CPS) can play a key role in the near future. These turbines are the prospective next development step within wind energy sector (compare Fig. 1).

2 Potential in Wind Energy

Although the development of wind energy technology is substantial, there is still great potential for improvements in the future. Figure 3 illustrates different levels of wind energy in the power grid and the integration to the society. The chart is divided in wind energy technology and community interests. Potentials exist on each level and CPS can be applied to take advantage of these potentials. The arrows symbolize the exchange of information between different levels, as communication is essential for CPS systems.

On the left hand side of Fig. 3, the base level consists of the wind turbine including technical specifications. Here, the potential from the manufacturing process up to the operation of a single wind turbine is considered. One level higher, the potential of wind turbines operating in a farm is illustrated. The main potential is an improved communication between single wind turbines resulting in an advanced forecast and an improved availability. The top level describes the interaction between wind energy and the main grid. Due to changing wind conditions, the power supply through wind turbines is challenging. Therefore, more intelligent and flexible grids like local smart grids, which are already installed, offer great potential in the future.

On the top level, the grid communicates with the society to enable a balance of demand and supply. There is a great potential in the integration of wind energy within the different levels of community interests. If the acceptance of wind energy in the society and particularly in the communes and families is increased, the energy transition with wind energy as a main player can be supported.

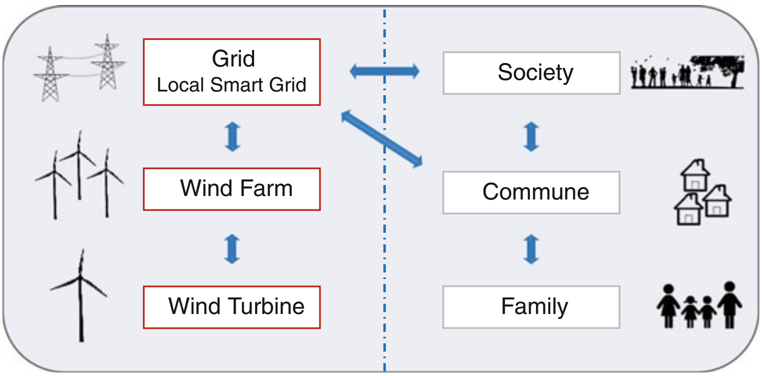


Fig. 3 Layers of wind energy in the power supply system related to the society

2.1 Wind Turbine

In wind turbines the total costs can roughly be divided into investment and operation and maintenance (O&M) costs. The proportion of cost is dependent on many factors, such as site location and country, weather conditions or political decisions. An onshore wind turbine in Germany, for instance, features a share of approximately 60 % investment and 40 % O&M expenses for an expected turbine's life time of 20 years [10].

One solution to mitigate the investment costs is to improve the manufacturing process of wind turbine components, such as rotor blades or foundations. Currently, most wind turbines worldwide are single-part to small batch series productions [11]. A standardized and automated run production of these components can mitigate investment expenses. Moreover, the quality of single components can be improved and O&M costs can be reduced. In this regard, CPS applied to production engineering is suitable to make use of the available potential.

One major challenge in wind energy are unsteady and rapidly changing weather conditions, such as wind direction or the intensity of wind gusts. Not only do weather conditions affect the efficiency and power supply of wind turbines but also the technical conditions. Early failure detection is a key solution to prevent avoidable subsequent damage of the turbine's components [3]. Moreover, maintenance work could be optimized by improved coordination and planning. Due to challenging accessibility, this is of significance especially for offshore plants.

Improved weather forecasts, accompanied by early failure detection and optimized maintenance planning are the main potentials for prospective wind energy developments in order to reduce O&M expenses. The development could be realized through the application of CPS by more intelligently working wind turbines that use measured physical data to automatically adjust to changing weather conditions. Additionally, automated control and supervision of wind turbine's condition could be highly improved with an advanced CPS.

2.2 Wind Farm

Currently, wind turbines operate autonomously within a wind farm [3]. They are connected through an internal wiring in order to feed the electricity into the grid but there is still no communication between the single wind turbines. The main potential in future wind farms is to develop an intelligent and automated communication network so that turbines can exchange valuable information. These information might be a measured extreme wind gust in the first row of wind turbines in a farm, which is then forwarded to turbines located in the back of the first row. This way these turbines are able to preventatively adjust their rotor blades and, therefore, avoid possible damage. Additionally, not only the communication of weather conditions within a single wind farm is desirable but also the exchange with neighboring wind turbines and farms.

Thus, local weather forecasts can be enhanced and consequently, the technical condition and also power supply of wind turbines will become more predictable.

As wind turbines within one farm operate under nearly the same weather conditions, the exchange of detailed generator system parameters for power production is another potential for reducing O&M costs. With this information, a turbine is able to detect a malfunction by using the monitoring information of other turbines [12]. In case of a deviation, the turbine can stop its operation and send a report to the control center.

These are possible solutions for a new generation of wind farms which offer a high reliability and performance level through the exchange of information within wind farms. The application of a CPS within wind parks makes the usage of these potentials feasible.

2.3 Grid and Local Smart Grid

One layer above the wind farm, the integration of wind energy to the power grid is considered. This layer has an interface to the non-technical aspects on the right hand side of Fig. 3, accounting for community interests. Here the main objective is a reliable supply of energy to cover the demand. One major challenge for the current power grid is the fluctuating power supply by renewable energies. As these already feature an inherent part of the power supply, grid stability concerns arise. To address this issue, supply and demand forecasts need to be improved. This can be realized through the implementation of an advanced CPS which operates with real-time sensed data and uses modelling algorithms and weather forecasting to predict the future demand and power supply of renewable energies [13]. Additionally, historical data could be used to detect load patterns and improve predictability.

Local smart grids are already applied in order to integrate wind energy into the power grid.

Generally, smart grids include a variety of operational energy measures that enable the system to control the production and distribution of energy. With smart grids it is feasible to process information from electricity generators and consumers in real-time for distributed power supply systems. They also include the availability of storages and support demand side management in order to forecast the power demand. Hence, these systems offer many benefits, such as a more efficiently operating power grid or improved coordination of energy sources [14]. In the future it is desirable to implement an advanced smart grid by means of CPS, which will be extended from local application to the entire grid. This system could coordinate and optimize the renewable energy share as part of the overall power supply on a national or even international scale, including every available energy source. Moreover, availability of storage or up to date electricity prices could be considered.

2.4 Community Interests

Community interests aggregates the layers society, commune and family. Aside from meeting the power demand, further community interests need to be considered as wind power plants are continuously extended. Due to limited space it is obvious that more and more turbines will be built closer to residential areas [15]. This fact increases the potential for conflicts about visual appearance, shade or noise generated through wind turbines. In this context it is important to increase the acceptance of wind energy within society. Hence, it is necessary to work together with local communities and even families. Local residents are supposed to be closely involved in wind energy projects at early planning stages, if desired even with financial participation. In fact communication between the different layers in Fig. 3 needs to be improved.

Application of CPS could contribute to an enhancement of communication amongst the wind lobby and the society and increase acceptance of wind energy. With advanced sensing technology in wind turbines, automated shutdown or adapted operation of the turbine is feasible. Thus, turbines could mitigate disadvantageous effects during relevant periods. Even applications for smart phones or tablets are conceivable which would allow the local residents to control single wind turbines in a way that reduces, for instance, the shade or noise of the turbine at certain times.

3 CPS in Wind Energy

Future potentials in wind energy technology are manifold but they share a common ground. Existing data of wind turbines and farms needs to be analyzed and new available data through advanced measuring technologies or communication networks needs to be generated in order to take advantage of the full potential. Therefore, the implementation of CPS within wind energy features great potential for future wind applications, as CPS builds a tight coupling between the digital and the physical world. All processes in wind energy from the accurate condition monitoring of wind turbine components in real-time to cost and time effective maintenance work could be controlled and optimized automatically.

CPS are already applied in various other areas, such as aerospace, automotive, robotics and healthcare to use the strength of both, the cyber and the physical world and enable novel and automated solutions in real time [16]. The implementation of CPS in renewable energy sources like wind power can support the integration into the energy system and increase efficiency, adaptability and reliability. Some application forms of CPS already exist in current wind turbine technologies and will be presented here.

3.1 Current Applications of CPS in Wind Energy

3.1.1 Condition Monitoring System

A Condition Monitoring System (CMS) is an engineered system which allows permanent monitoring of the machine condition by measuring and analyzing physical values, such as temperature and vibration [17]. It is a reliable and fast-reacting security system for detecting and predicting mechanical and electrical faults in main machine components, such as roller bearings, gears, blades or the induction machine. CMS can operate as control systems which modify the operation mode in case of suspicious measured data or emergencies.

The data is measured through integrated sensors in real time. Overall, the requirements for sensor systems within a CMS are very high. The sensors need a durability of more than 20 years and should neither be influenced by any kind of load nor by environmental or operational conditions. An example for an advanced sensor technology for wind turbine application are fiber Bragg grating sensors (FBG) which are able to measure temperatures, pressures and vibrations [18]. FBG sensors feature fast responses and are suitable for monitoring systems due to their small size. In current wind turbine applications, they are used for blade supervision, offering high reliability and durability.

One common application of CMS are vibration analysis to detect any deviation of the normal stress range [19]. CMS gathers high frequent data from vibration sensors of the wind turbine's drivetrain at different positions. For example, the V112 wind turbine from Vestas features 13 sensors on the drivetrain, where six of them are located in the gearbox [20]. The other sensors measure the speed and characteristics of the main bearing, nacelle and generator. The time history response is then measured and can be converted to the frequency domain by Fourier transformation. With statistical algorithms the damage cases can be predicted at an early state and damage to wind turbine components can be avoided. In addition, alarms are classified according to the severity and lead times for each alarm. Also the related maintenance work that has to be performed is available. However, state of the art CMS in wind turbines collect data in discrete time intervals and monitor only some selected measurement points. Hence, it is desirable to achieve a CMS in the future, which is capable to measure and monitor continuous loads of all relevant machine components. In this regard, a change from reactive to predictive maintenance is essential in order to enhance the O&M of wind turbines in the future.

A statistical evaluation of wind turbines operating in Germany showed that frequent failures are arising in electrical and plant control systems [21]. Furthermore, the amount of failures in offshore turbines is much higher than for onshore sites. The higher velocity of wind gusts at sea has a bearing on the turbine components, such as the rotor shaft and the gearbox, leading to higher forces and stresses. However, the important aspects that need to be considered are downtime and cost factors resulting from a damage event. The maintenance time for a gearbox, drivetrain or generator is much higher than for other components of the wind turbine due to the complexity of these parts. Hence, longer downtimes are expected. Preventing downtime is even more important for offshore wind turbines due to deficient accessibility.

The gearbox has by far the highest cost per failure with average maintenance cost of 230,000 € and a maintenance time of about 200 h [22]. A downtime resulting from a gearbox damage, for instance, leads to a production blackout for at least one week in onshore and even more in offshore turbines.

This emphasizes the importance of a functional CMS, applied in wind turbines to reduce the amount of major failures. The main benefit of CMS implementation is the opportunity to monitor all significant components and intervene before any damage appears. Therefore, application of CMS is particularly worthwhile in expensive and sensitive machine components like the gearbox.

3.1.2 Supervisory Control and Data Acquisition

The physical environment of wind turbines is the most important factor concerning the operation status and power output. For example, the wind turbine performance increases with third power of wind speed, so an incorrect forecast of wind characteristics can result in significant deviation of the gained power [23]. Therefore, operators of wind turbines and farms are interested to get an overview of all relevant operating conditions. These are namely wind properties, like wind speed, turbulence and direction, energy related information, such as power output or generator torque, and vibration or temperature data [24]. The manifold and required wind turbine data can be provided through Supervisory Control and Data Acquisition (SCADA), an accurate and modern sensing technology system. SCADA systems are able to collect different data streams at discrete time intervals through sensors and transfer this information to a supervisory control center. An overview of the wind turbines condition up to nearly real time is then provided for the operator and remote control of the physical system is possible. The Information is stored online and available to operators and customers from all over the world. Hence, SCADA systems facilitate automated remote monitoring and are capable of optimizing wind turbine operation.

The system builds a human-to-machine interface and provides an opportunity to reduce O&M expenses. According to [25], the most important characteristics of SCADA systems can be summarized as follows:

1. Monitoring
   Concentrated operating data is displayed and nearly real time monitoring is possible. A wind farm management system illustrates the current operation status of every wind turbine.
2. Reporting
   The SCADA System allows a database-driven performance analysis with access to relevant information from all over the world.

3. Control
   SCADA offers remote control to easily change the wind turbine's mode of operation from the control center, like start or emergency stop events.

An exemplary application of SCADA systems is the ice detection on rotor blades. Icing reduces the efficiency of wind turbines as the aerodynamic is negatively influenced. In addition, ice dropping poses potential danger to neighboring wind turbines or even buildings and people. There are already solutions to prevent icing by using a numerical weather prediction model that determines atmospheric conditions necessary for ice building. The system identifies potential danger for icing and triggers an alarm [26].

Moreover, solutions with integrated sensors and heating elements in the rotor blades are available. The sensors capture the ambient condition and activate the heating element in case of icing [27].

Although the current performance of SCADA systems is a substantial progress in wind energy technology, further improvement of these systems is necessary. In the future, continuous measurements of system parameters are worthwhile in order to optimize the O&M of wind turbines and farms. Moreover, to take advantage of synergy effects, a close communication among CMS and SCADA systems is required, as these systems currently operate autonomously.

3.1.3 Smart Grid and the Integration of Wind Energy

Not only does renewable energy production help to add green energy to the grid, but it also brings a significant degree of variability along. Since the contribution of renewable energy supply to the grid is continuously increasing, the grid is faced with the challenge to effectively integrate different sources. Especially wind power characteristics are highly variable over multiple patterns of time. The wind turbine performance is affected by seasonal and diurnal changing weather conditions. Consequently, it is crucial to have a power grid that is capable of dealing with the intermittent energy supply of renewables.

Power grids that intelligently integrate suppliers and consumers to efficiently deliver electricity are called smart grids [28]. These types of cyber-physical systems can promote the substitution of conventional power plants by distributed renewable energies, as information of energy demand and consumption as well as availability of storages are considered at the same time. Furthermore, smart grids feature real time pricing and solutions for consumers to obtain electricity only when it is available from renewable resources. Voltage regulation and load following as it can be realized by smart grids can reduce operation expenses based on marginal production costs [29].

All over the world, one can find successful projects where smart grid structures were implemented in order to make energy management more efficient. In 2011 for example, Enel installed the first European smart grid system in Isernia, Italy. Today, this pilot project enables a bi-directional flow of electricity generated from renewable energies on low and medium-voltage power grids. Some features of the system are sensors for advanced grid monitoring, storage in form of lithium-ion battery technology, a recharging station for electric vehicles and monitoring equipment that can be used to control residential consumption [30].

Another example for a smart grid system can be found in Mannheim, Germany. Mannheim is known as the "Smart City Mannheim", since it is one of several research projects that were set up under the German government funded program "E-Energy". Within this project, the power grid of the future is tested. Consequently, it demonstrates how renewable energies can be optimally integrated into the grid. With the help of a broadband powerline, the grid was turned into a real time communication platform which is connected to all households, decentralized suppliers and measuring devices. Overall, 3000 households will be equipped with devices that enable the control of household activities and the communication with other smart grid participants [31].

Even though there are multiple examples for successful smart grid implementations, there are still issues that need to be faced. When it comes to smart grids one of the most critical aspects are the questions of data ownership and access among the utilities of the various energy sources, as limited information is available in public. Furthermore, grid security needs to be considered under completely new aspects. Another complex subject is to resolve the question of the smart grid operation guidance. The role of the individual components, such as wind energy, in smart grids needs to be defined exactly in order to keep the system´s power quality, reliability and cost properties. Therefore, standards need to be defined, in order to enable proper communication between the individual parts [14].

In summary, smart grid technologies in combination with supporting policies and regulations are essential for the energy transition. By supporting a decentralized power supply, a smooth integration of high shares of variable renewable energies is feasible. Smart grids also integrate consumer interests and provide flexibility on the demand side. In future smart grids it will be necessary to temporarily decouple the power demand from the supply and satisfy the demand in particular during times of high shares of renewables. In this regard, detailed consumer load information of households and industry are essential. The processing of this data is feasible through an advanced CPS in the future, which enables smart grids on a national or even international scale.

3.1.4 Interaction of Wind Energy and Storage Systems

Electricity storage is extremely useful for adding more flexibility into the power grid. This is especially helpful with regard to the increasing implementation of wind turbines and farms into the grid, showing a broad range of variable characteristics.

Generally, two types of energy storing systems can be distinguished [32]. On the one hand, centralized bulk storage systems such as gas tanks and pipelines can provide power over a long period of time. Compared to bulk systems, the power output in distributed storage systems is smaller. Examples for distributed storage technologies include lithium-ion batteries, lead acid batteries, types of flow batteries, thermal storage systems, flywheels, supercapacitors, and hydrogen storage [33].

Depending on the energy demand, the storage duration and the amount of energy that should be stored, it needs to be evaluated whether centralized or distributed storage systems are appropriate to use.

Currently, one of the most critical aspects about distributed storage are the expenses. Even though the technologies are available, the deployment of distributed energy storage on a larger scale is still too expensive. Independent from the cost issue, multiple benefits can be provided by distributed storage systems. Their implementation to the grid can help to stabilize the system and to enhance the power quality by applying frequency and voltage regulations. Additionally, the variability caused by renewable energy sources within the grid can be smoothed. Any kind of peaks, either resulting from renewable energies or short term loads, can be shaved by the use of distributed storage systems. They can operate as a back-up power source to provide power on a short term basis, while a distributed generation source comes online in an islanded micro-grid [32].

Furthermore, distributed storage systems in combination with CPS enable smart storage, so that the interaction of consumers and suppliers can be optimized. An example for smart storage are electric vehicles that are able to feed electricity back into the grid during peak times. Services like ramp-smoothing, voltage control, and frequency regulation could be provided by the intelligent control of the charging behavior of an electric vehicle. Since electric vehicles are typically charged at night, they are qualified to absorb power from wind generation, which tends to peak often during these times [32]. However, the technology for feeding electricity back into the grid is still in development, but it will not increase the cost of an electric vehicle charger significantly.

Within the variety of distributed storage technology, lithium-ion batteries are receiving the most attention in current research and pilot projects. General Electric for example integrated batteries in their wind turbines in California in order to allow short term energy storage. Consequently, ramp control, predictable power and frequency regulation are enabled [34].

Overall, the main factors preventing wider deployment of distributed energy storage at the current state are high costs and low life expectancy of distributed storage systems. However, the implementation of distributed storage systems in CPS offers great benefits and potential, especially in combination with renewable energy generation. Furthermore, the expenses for distribution and storage are continuously decreasing, so one can assume that distributed storage systems may play a significant role in the future power grid.

3.2 *Future CPS in Wind Energy*

This chapter will give an outlook of how CPS and wind energy could be combined and benefit from each other in the future. Based on already existing forms of CPS in wind energy, an advanced adapted CPS will be able to use the strength of both technologies and make use of the still available great potential of wind energy. These are mainly the smooth integration of wind turbines to the power grid and the optimization of operation and maintenance achieving a more reliable and economic system.

A proposed future CPS in wind energy is illustrated in Fig. 4. It considers all important variables and factors related to wind energy. An overview of all correlations as well as data and electricity streams is presented. The left hand side of Fig. 4 represents the physical system, focusing on the integration of wind energy to the supply system. On the right hand side, information such as the collected operation characteristics, wind data or the power demand is collected within the cyber system. One step further, this information is used to analyze, optimize and control all processes, for example the optimized implementation of wind energy into the grid, in one central unit.

3.2.1 Wind Turbine/Farm

The technology of wind turbines is illustrated through the wind turbine/farm unit, which features two outputs. One stream is connected to the grid providing households and storages with electricity, the other stream illustrates the high data

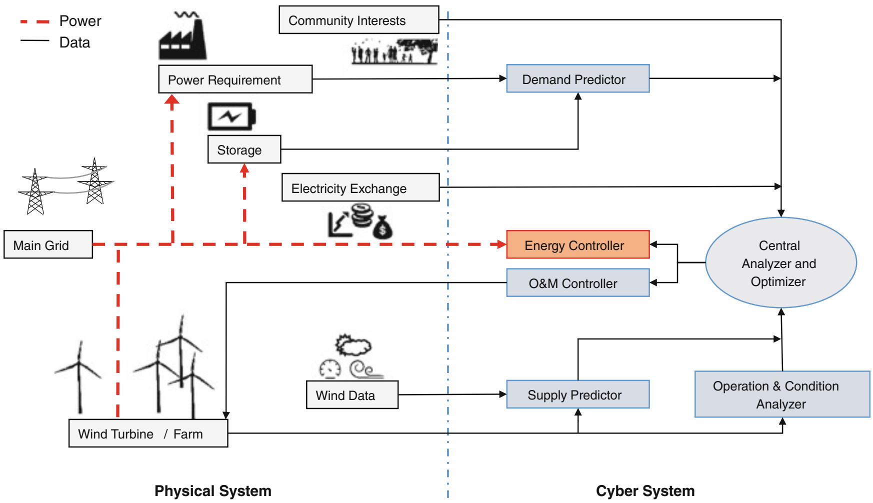


Fig. 4 Future CPS in wind energy

volume generated by CMS and SCADA information. This data flow is directly coupled with the cyber system units, which process this valuable information.

In order to enhance the quality of CMS and SCADA systems, the accuracy and reliability of the sensor system need to be improved. For example, an inaccuracy in the measurement sensors could result in fault interpretation and action on the control side. In the worst case, this might even entail damage of components or system downtime.

The challenge to achieve a more reliable response in future CPS applications can be addressed on the hardware side by increasing the number of sensors. These can be connected in a series of nodes combined with a fast calculating processor. The high amount of input then leads to more reliable and concentrated output signals.

Additionally, the sensor quality needs to be improved. With the application of intelligent sensors that measure temperatures, vibrations, pressures and acoustic emissions in combination with dynamic and steady state signals of the turbine controller, system component failure can be detected at early stages before any serious damage can occur [35]. Through acoustic emissions, for instance, mechanical dysfunctions, such as bearing damages, rotor imbalance and flow instabilities can be identified and processed. Optical fiber sensors meet the requirements of wind turbine sensors best, since they have fast response times and a high durability of over 20 years.

Besides an increase of the reliability it is also desirable to reduce the monitoring system's dependability. This could be achieved in the future CPS in wind energy through the development of new applications such as cameras, alarm systems or drones. A malfunctioning sensor is then substituted by a different sensor measuring the same kind of system behavior.

With the implementation of network cameras it is conceivable to monitor the health condition of wind turbine components. For instance, many existing applications use thermographic methods, like an infrared camera, to visualize a variation in gearbox temperature [36]. Moreover, detection of delamination, subsurface cracks or wrinkles in the fiberglass of the rotor blades are possible. Those failures could have significant consequences for the system and are difficult to identify though conventional maintenance. The method is already applied in offline mode but in the near future it is worthwhile to integrate it in wind turbines to reduce the dependability within the CMS.

Another example is the utilization of drones for maintenance work, which is an exciting innovation for the wind energy industry. There are several companies that do research work on implementing this technology in order to reduce O&M costs. This is especially significant for offshore wind farms where accessibility is challenging. Due to the latest efforts in research on drone technology for non-military applications, the inspection of wind turbines with the help of drones has proven to be very useful. A drone inspection would offer higher resolution at less costs than a normal maintenance service [37].

It is conceivable to investigate several wind turbines in one process. An infrared camera installation would also be possible. Currently, there are still some issues which need to be taken care of. These concern the evaluation and the analysis of the provided pictures and the ability of drones to fly autonomously while keeping a certain distance to the rotor blades. However, one can assume that drones will play an important role in the near future of wind energy. The application is well suited for the proposed CPS and features the potential to significantly reduce the O&M costs of wind turbines and farms.

A different approach to control and monitor the operation of wind turbines are communication networks in wind farms. Currently, wind turbines operate autonomously in farms without any communication among each other, although several communication options are already available, such as fiber optics, Ethernet and wireless technologies [38].

Hence, it is conceivable that wind turbines share weather related information in order to improve local wind forecasting. This information includes, for instance, wind direction and speed data received through SCADA systems. Therefore, wind turbines could anticipate the ideal operating position or extreme wind gusts and adjust the blade angles (pitch control) or the orientation of the nacelle (yaw control) in time.

Another way to take advantage of exchanged information among neighboring wind turbines in a farm is shared operating data. Assumed that the wind turbines are of the same kind, malfunctions could be detected by comparing the system condition to other turbines. For instance, a deviation in temperature or pressure could trigger an alarm in the control center.

In summary, the exchange of information between single wind turbines offers a great potential to increase the level of automation and reduce the maintenance costs. In the proposed future CPS in wind energy it is then further intended to link the autonomously working CMS and SCADA systems in order to combine the strength of both information in a central control center.

3.2.2 Wind Data

One of the main challenges of wind energy is to achieve improved weather forecasts in order to anticipate various load conditions and the power supply of wind turbines and farms. Moreover, the performance of the wind turbines controller is dependent on the available wind information [39]. Hence, extended and optimized wind data analysis is necessary. Within the future CPS system, larger amounts of data can be measured and evaluated centrally. This includes both, historical and current data.

Analysis of historical data is especially relevant for the purpose of site selection, which has a significant influence on the potential energy yield of the wind turbine. A well-chosen site entails more cost-efficient wind energy. The high amount of historical data could be managed and analyzed with an advanced CPS.

To improve up-to-date weather forecasts, not only advanced measurement methods but also a better communication network among existing weather data acquisition systems are crucial. An example for modern measurement tools are Light Detection and Ranging (LiDAR) systems, which send laser pulse signals to the environment and measure the reflection. These systems can accurately determine wind speed and direction and feature marginal fault liability [40]. The advantage over conventional measurement methods, such as a measurement mast, is the high flexibility provided by this technology. Hence, LiDAR is also suitable for mobile applications. The current distribution of LiDAR systems is expandable. However, systems like LiDAR could play an important role within the future CPS in wind energy to accurately and effectively predict wind conditions.

Another solution to improve weather forecasts is a closer collaboration between local and independently working weather stations. For instance, weather information from aviation or navy could be combined and integrated to conventional weather models. This kind of weather communication network could improve local wind turbine site forecasts.

3.2.3 Supply Predictor

Coming along with more sophisticated and improved weather models, the supply of energy, can be predicted more accurately. This is illustrated by the supply predictor unit in Fig. 4. Especially when it comes to wind energy, detailed knowledge about the current and future weather conditions are crucial for minimizing the impact of significant weather events on the energy supply. Due to more accurate and real time communication of weather data, predictions of wind power supply would be based on shorter terms and it would be possible to predict concrete provision rates on an hourly basis. Additionally, the information of CMS and SCADA is used to consider wind turbine availabilities.

Overall, the further implementation of CPS in wind energy will without a doubt lead to significant changes within the energy generating field. The power supply prediction for wind turbines will be improved considerably through the growing presence of CPS.

3.2.4 Operation and Condition Analyzer

In the operation and condition analyzer, the information of single wind turbines as well as the one from entire wind farms is collected and analyzed. For example, vibration or oil analysis are methods that are used to gather information about the system's current condition.

This information can be used to detect possible failures and critical elements of the system in time. It is crucial that sufficient warning time for upcoming maintenance is provided by the monitoring systems.

Especially the implementation of a service life calculation in the monitoring system is desirable to detect potential system weaknesses in advance.

Predictive maintenance uses continuously collected data to simulate the transient load distribution of critical elements. From the generated load patterns, the operational live span of the system can be derived online and thus maintenance services can be planned ahead accordingly. When using predictive maintenance, maintenance on machines is only performed when it is required. Unexpected downtime of the system can be reduced almost entirely. This results in remarkable cost savings, since not only the maintenance time is minimized, but also the amount of used spare parts and supplies [41].

This could lead to a more sophisticated and reliable condition monitoring system.

3.2.5 Demand Predictor/Power Requirement

Considering the demand side of Fig. 4, accurate power requirement forecasts are crucial within the future CPS in wind energy. At the current state, repetitive patterns in energy consumption can be observed. In Germany for example, there is an energy demand peak during lunchtime within the week and the energy demand is significantly lower during the weekend compared to working days [42]. Furthermore, on an annual basis the energy consumption is higher in winter than in summer. These correlations can be used in the demand predictor receiving information from the power requirement unit, illustrated in Fig. 4.

Nowadays, these repetitive variations are already well integrated in the demand forecasting.

However, with the implementation of intelligent communicating systems, extraordinary demands could be estimated more accurately. An example is an increased demand during the half time of soccer games, anticipated by a CPS that includes the level of viewers.

Furthermore, the load pattern fluctuation could be controlled and used by intelligent operating devices that monitor the current demand. Depending on the load, these systems could charge or activate themselves when it is energetically smart to do so. For instance, this could be during the night or at the weekend.

Through the active communication of load devices with other components of the CPS, forecasts of demand will become more efficient. For example, predictable energy consumption, like smart dishwasher waiting for low domestic energy demands to turn on, can be considered in advance. Therefore, demand responsive load management will be feasible.

3.2.6 Storage

Another factor which is considered in the demand predictor unit in Fig. 4 are storage technologies. The storage of energy, in particular wind energy, provides a great potential in the proposed future CPS with wind energy.

A combined system of wind energy and storage technologies is illustrated in Fig. 5.

The figure shows solutions to storage excess of wind energy on a short and long term basis. For the short term solution, the implementation of batteries in wind

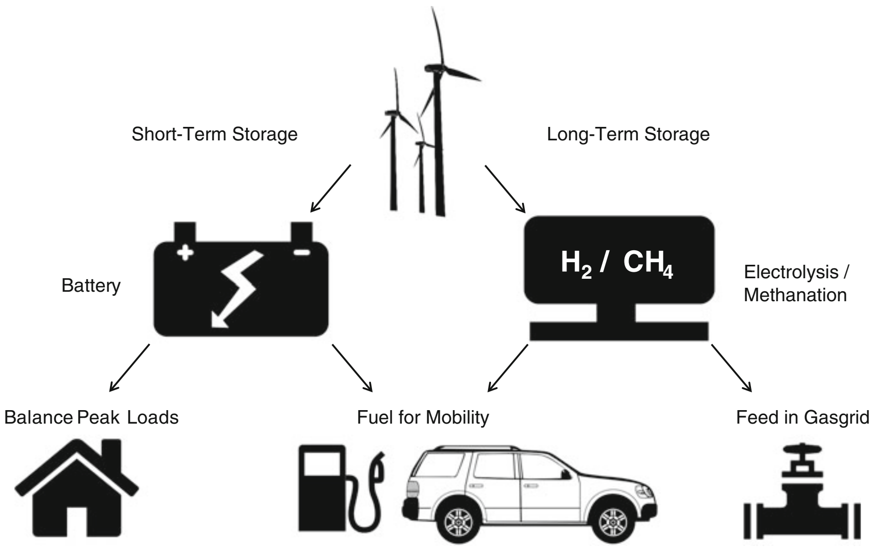


Fig. 5 Combined system of wind energy and storage technologies

turbines is presented. This enhances the turbine's flexibility as frequency regulation is feasible [34]. Therefore, unpredictable changes of wind conditions and load demand can be leveled out. Moreover, this enables excellent features to balance the power grid during peak loads [43, 44]. Another possible short term application are electric cars, which is a growing market. With intelligently charging electrical vehicles which are capable to feedback power in the grid during peak times, ramp smoothing, voltage control and frequency regulation could be provided.

For long term solutions, pump storages feature very high efficiencies [45]. Unfortunately, the development of this technology is limited through geographical conditions. Hence, the Power-to-Gas method is presented here as it offers high potential for future applications [46, 47]. Hydrogen can be produced through electrolysis and used for hybrid cars or stored in the existing gas grid. With the production of methane through methanation, a post process of the electrolysis, an even higher energy density is feasible.

In general, smart storage of energy can be enabled through the intelligent communication between energy consuming and energy generating elements within the power grid.

3.2.7 Electricity Exchange

Electricity exchange is another element considered in future CPS in wind energy. This unit section will include information required to optimize the system from a financial point of view.

The proposed CPS is supposed to gain experience through historical and current price developments, but also political decisions affecting the energy system are significant. The bearing of political decisions on price trends is exemplarily illustrated through the recent market situation of wind energy in Germany. In order to reach the high aims of the energy transition, the German government sets price incentives for renewable electricity. Renewable energy which is directly sold at the electricity exchange obtains financial support. The support is financed by the domestic electricity price, the so called “EEG-Umlage” [48].

This reform has definitely affected the energy price on a short- and a long term basis. Therefore, political decisions play an important role within the electricity exchange unit.

3.2.8 Community Interests

Besides financial and demand concerns, community interests play an important role in the future CPS. Increased acceptance of wind energy within society supports the integration of this technology and, moreover, helps to find new locations for further wind turbines or farms. Hence, community interests are considered while making decisions in the central analyzer and optimizer block.

Some of the mostly discussed issues when it comes to the effect of wind energy on society are shading vorticities, noise, icing, influence on land- and cityscape, nature protection considerations or the influence on navy and aviation. These aspects have to be faced, especially since the current development of wind energy will continue in the future and wind turbines as well as households will be located closer to each other. Nowadays, systems that address these issues and increase public acceptance are already existing. Senvion, for instance, developed a sound management system to reduce noise emissions at certain time periods or wind directions [49]. They also apply a shadow management system, in which the wind turbine automatically avoids shading vorticities if required. These are some exemplary endeavors of wind turbine manufacturers to consider community interests. However, broader measures are required to further increase acceptance in the future. This could be the implementation of apps within smartphones or tablets to allow the residents to influence the wind turbine operation in particular time periods.

3.2.9 Central Analyzer and Optimizer

The central analyzer and optimizer is the main unit of the proposed future CPS in wind energy. In this unit, the large data volume is merged. In particular, information from the supply and demand predictor, data from the operation and condition analyzer, the community interests and the electricity exchange data are linked. The valuable data stream is then stored and archived in order to process the information and access stored data subsequently. Thus, the central analyzer and optimizer can gain experience and become a self-learning and self-optimizing system.

The essential function of this unit is to analyze and optimize all processes based on all input parameters. The considered processes mainly include the power distribution as well as the operation and maintenance optimization of wind turbines. Application of such an advanced CPS combined with wind energy would be able to upgrade wind energy development to the next level, since every possible influencing factor is simultaneously considered in one system. The automated and optimized decisions generated in the central analyzer and optimizer unit are then executed through the energy and the O&M controller.

3.2.10 Operation and Maintenance Controller

Within the O&M controller, the entire wind turbine and farm operation as well as the maintenance is controlled. These areas feature great potential to increase efficiency and reduce O&M expenses, which can be utilized in the future CPS with wind energy.

Concerning the operation of a single wind turbine, the current system condition plays a key role when decisions have to be made in the central analyzer and optimizer. One example are measured system abnormalities through CMS which do not entail an emergency stop or maintenance necessity but will negatively affect the service life of the turbine. Consequently, the operating time of this particular wind turbine could automatically be reduced in order to avoid damage of components or maintenance operation. The reduced supply of the considered turbine is then supposed to be delivered by other turbines within the farm or energy grid. This implementation is feasible in the proposed CPS, where all important parameters are considered. Besides the system condition, the energy demand and electricity price are also included. In the future CPS it is then conceivable that every single turbine automatically identifies and adapts its optimum operation mode considering these input parameters. The optimum operation mode could also be applied to specific wind farm characteristics. For instance, wind turbines suffering from wake effects of neighboring turbines could be less included to the power supply than wind turbines operating at their optimum. This would decrease the frequency of undesirable operating conditions and extend the service life.

Further considerations affecting the operation of wind turbines and farms in the proposed CPS are community interests. Disturbing factors generated through wind turbines need to be measured and documented and be included into the decision making process within the central analyzer and optimizer unit.

With respect to the maintenance of wind turbines, the future CPS in wind energy is capable of coordinating and optimizing all processes. The considered influencing factors for decision making are manifold and include detailed component conditions measured through CMS as well as the availability of maintenance equipment and workers. Solutions to lower O&M costs are highly available in this kind of

CPS. For instance, maintenance work would only be performed in times of appropriate weather conditions in order to reduce the maintenance duration and, consequently, systems downtime. Moreover, the maintenance operations could be concentrated so that several wind turbines can be serviced at the same time. Another way to decrease O&M expenses is to consider the entire energy market and the electricity exchange. Maintenance would then only be carried out during time periods when the profit of the wind turbine is expected to be low.

3.2.11 Energy Controller

The other controller of the proposed CPS in wind energy is the energy controller which is responsible for the energy distribution within the power grid. In this unit, the entire energy system including all available sources and storages that are needed to supply the required demand is considered as an advanced smart grid. The controlling is then based on the optimized decisions resulting from the central analyzer and optimizer unit.

The current situation of energy control can be illustrated exemplarily through the situation in Germany. There, supply security is not only provided by a leveled out mixture of energy sources, but also through a networking, highly reliable power grid [50]. Generally, the transmission grid operator is responsible for the reliable operation of the power grid. Therefore, the power balance method is used. It compares the available power plant capacities with the peak load of the year. If necessary, energy is imported in order to keep the supply steady. Furthermore, the load for the next three years is estimated with the help of historical data.

However, the described method only considers the power demands in Germany. This is critical, since the European internal electricity cross-border market is constantly growing.

Energy exchange over borders and long-distance balancing strategies become more and more important. Especially within the context of a steadily growing number of renewable energy generating power plants and inherent varying energy supply.

Additionally, the decentralized power generation that comes along with the implementation of renewable energy supplier in the grid brings along the risk of uncontrollability.

The grid operators cannot control all the low-capacity, distributed resources, so it might be the case that the grid is not balanced. Therefore, appropriate forms of smart control features are necessary [50].

This could be realized through the advanced CPS which considers all significant influencing factors. Hence, the optimal solution for energy controlling can always be found on a national or even international scale. The grid operators would obtain the ideal tool to achieve a well-balanced power grid in the future.

4 Outlook and Conclusions

It was shown that wind energy has been established as the major renewable energy source in the current power grid. This technology is essential to accomplish the ambitious climate targets set by the governments and expedite the energy transition. Wind energy is still a growing market and features high potential on different levels. These are mainly the improvements of wind turbine or wind farm operation but also the integration to the power grid through an advanced smart grid and consideration of community interests is substantial.

The future potential in wind energy is multifaceted but shares a common ground. Current and prospective available data has to be analyzed and optimized in order to take advantage of the entire potential. In this regard it has been demonstrated that a CPS can be a powerful approach to further develop the wind energy sector. The proposed future CPS considers significant influencing parameters on wind energy technology and the integration to the main power grid. It also includes current forms of CPS, like CMS and SCADA, and is capable of automatically controlling and supervising all processes. Consequently, the operation of wind turbines and farms can be optimized and efficiency be increased. Furthermore, lower O&M costs are feasible.

The implementation of the proposed future CPS in wind energy, which operates reliably, efficiently and in real-time can contribute to a more stable and robust power grid in the future. The system would consider the intermittent energy supply by wind turbines and simultaneously take the condition of every single wind turbine and component into account. Hence, the time has come to take the next step and transport wind energy development to a higher level by making use of CPS for wind energy.

References

1. Office of Energy Efficiency and Renewable Energy. http://energy.gov/eere/wind/history-wind-energy. Accessed 30 Nov 2015
2. Erneuerbare Energien. http://www.erneuerbare-energie-windenergie.de/windenergie-deutschland. Accessed 30 Nov 2015
3. Hau E (2008) Windkraftanlagen. Grundlagen, Technik, Einsatz, Wirtschaftlichkeit. 4. Auflage, Springer, Berlin
4. Bundesverband für Wind. BWE. https://www.wind-energie.de/. Accessed 30 Nov 2015
5. Gonzales JS, Arantegui L (2014) JRC wind status report—technology, market and economic aspects of wind energy in Europe. Luxembourg: European Commission, 2015
6. Global Wind Energy Council. GWEC. http://www.gwec.net/. Accessed 30 Nov 2015
7. Blauer-Engel-Preis 2015, Umwelt Bundesamt. https://www.umweltbundesamt.de/. Accessed 30 Nov 2015
8. Bundesministerium für Wirtschaft und Energie. Security of supply. Accessed 30 Nov 2015
9. Agentur für Erneuerbare Energien. Forschungsradar Energiewende: Stromgestehungskosten verschiedener Erzeugungstechnologien. http://www.forschungsradar.de/studiendatenbank.html. Accessed 30 Nov 2015

10. Wallasch AK, Lüers S, Rehfeld K (2013) Kostensituation der Windenergie an Land in Deutschland. Deutsche WindGuard GmbH, Varel
11. Heumer W (2012) Modulbaukasten soll Kosten für Offshore-Windenergieanlagen senken. s.l.: VDI
12. Young K, Ahmed M (2014) Communication network architectures for smart-wind power farms
13. Luo T, Tham CK (2013) Sensing-driven energy purchasing in smart grid cyber-physical system. IEEE Trans Syst Man Cybern 43:773–784
14. Kempener R, Komor P, Hoke A (2013) Smart grid and renewables—a guide for effective deployment. s.l.: International Renewable Energy Agency
15. Devine-Wright P (2005) Beyond NIMBYism: towards an integrated framework for understanding public perceptions of wind energy
16. Rajkumar R, Lee I, Sha L, Stankovic J (2010) Cyber-physical systems: the next computing revolutio,. Anaheim, CA
17. Amirat Y, Benbouzid M, Al-Ahmar E, Bensaker B (2009) A brief status on condition monitoring and fault diagnosis in wind energy conversion systems
18. Micron Optics. Optical fiber sensors guide, fundamentals and applications. http://www.micronoptics.com/wp-content/uploads/2015/06/Updated_Optical_Fiber_Sensors_Guide_130529.pdf. Accessed 30 Nov 2015
19. Carden P, Fanning EP (2008) Vibration based condition monitoring: a review. Pennsylvania State University: SAGE Publications
20. Robertson A (2012) Monitoring the operation of wind turbines. In: Vestas Northern Europe Renewable Energy II Conference, Vaasa, Finland
21. Hahn B, Durstewitz M, Rohrig K (2006) Reliability of wind turbines—experiences of 15 years with 1,500 WTs. Institut für Solare Energieversorgungstechnik (ISET), Kassel
22. Caroll J, McDonald A, McMillan D (2015) Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines
23. National Wind Watch. https://www.wind-watch.org/faq-output.php. Accessed 30 Nov 2015
24. Li W, Kusiak W (2010) The prediction and diagnosis of wind turbine faults. Department of Mechanical and Industrial Engineering, 3131 Seamans Center, Elsevier, The University of Iowa, Iowa City
25. Senvion. Integration und Control, Ganzheitliches Management für Windenergieanlagen. https://www.senvion.com/. Accessed 30 Nov 2015
26. Vestas Cold Climate Solutions, Ensuring safe and efficient energy production in cold climate conditions. http://nozebra.ipapercms.dk/Vestas/Communication/Productbrochure/TurbineOptions/VestasColdClimateSolutions/. Accessed 30 Nov 2015
27. Nordex. Anti-Icing, Higher Yields in Icy Climates. http://www.nordex-online.com/fileadmin/MEDIA/Produktinfos/EN/Nordex_Anti-Icing_en.pdf. Accessed 30 Nov 2015
28. Gaviano A, Weber K, Dirmeier C. Challenges and integration of PV and wind energy facilities from a smart grid point
29. Clastres C (2011) Smart grids: another step towards competition, energy security and climate change objectives
30. Enel (2011) Enel Distribuzione: Italy’s first smart grid in Isernia, Rome
31. PPC AG. Moma-Smart City Mannheim. http://www.ppc-ag.de/268-1-Moma—Smart-City-Mannheim.html. Accessed 30 Nov 2015
32. Kempener R, Komor P, Hoke A (2013) Smart grid and renewables—a guide for effective deployment. International Renewable Energy Agency, s.l.
33. International Renewable Energy Agency (2012) Electricity Storage and Renewables for Island Power. s.l.: International Renewable Energy Agency
34. GE. Integrated Storage. https://renewables.gepower.com/wind-energy/technology/integrated-storage.html. Accessed 30 Nov 2015
35. Ciang C, Lee J, Bang H (2008) Structural health monitoring for a wind turbine system: a review of damage detection methods. Measur Sci Technol 19:1–20

36. García Márquez F, Tobias A, Pérez J, Papaelias M (2012) Condition monitoring of wind turbines: techniques and methods
37. Windpower Engineering Development. http://www.windpowerengineering.com/maintenance/heres-what-drones-might-do-for-the-wind-industry/. Accessed 30 Nov 2015
38. Hansen A, Soerensen P, Iov F, Blaabjerg F (2005) Centralised power control of wind farm with doubly fed induction generators. s.l.: Elsevier
39. Jassmann U, Berroth J, Matzke D, Schelenz R, Reiter M (2014) Model predictive control of a wind turbine modelled in Simpack. J Phys Conf Ser 524:012047
40. Frauenhofer IWES. http://www.energiesystemtechnik.iwes.fraunhofer.de/. Accessed 30 Nov 2015
41. Predictive Maintenance. https://www.maintenanceassistant.com/maintenance-triggers/. Accessed 14 Dec 2015
42. Sotiropoulos E (2012) Modeling of German electricity load for pricing of forward contracts. ETH Zürich, Zürich, PSL-1122
43. Wietschel M et al (eds) (2015) Energietechnologien der Zukunft. Erzeugung, Speicherung, Effizienz und Netze. Springer Vieweg, Heidelberg
44. Sterner M, Stadler I (2014) Energiespeicher—Bedarf, Technologien, Integration. Springer Vieweg, Heidelberg
45. Aebli M, Trüssel J (2012) Rentabilität von Pumpspeicherkraftwerken. Gruppenarbeit EEH Power Systems Laboratory, Eidgenössische Technische Hochschule (ETH), Zürich
46. Trost T, Horn S, Jentsch M, Sterner M (2012) Erneuerbares Methan: Analyse der CO2-Potenziale für Power-to-Gas Anlagen in Deutschland. Z Energiewirtsch 36:173–190
47. Nutzen der Power-to-Gas-Technologie zur Entlastung der 110-kV-Stromverteilungsnetze. http://www.dvgw-innovation.de/fileadmin/innovation/pdf/g3_03_12_erg.pdf. Accessed 22 Apr 2016
48. Energiewende direkt. Bundesministerium für Wirtschaft und Energie. https://www.bmwi-energiewende.de/EWD/Redaktion/Newsletter/2014/30/Meldung/strompreise-so-wird-die-eeg-umlage-festgelegt.html. Accessed 30 Nov 2015
49. Integration und Control, Ganzheitliches Management für Windenergieanlagen. https://www.senvion.com/global/de/wind-energy-solutions/operation-scada/. Accessed 22 Nov 2015
50. Bundesministerium für Wirtschaft und Energie. Security of Supply. https://www.bmwi.de/BMWi/Redaktion/PDF/Publikationen/Studien/leitstudie-strommarkt-2015,property=pdf,bereich=bmwi2012,sprache=de,rwb=true.pdf. Accessed 23 Nov 2015

Transfer Printing for Cyber-Manufacturing Systems

Varun Ravikumar, Ning Yi, Vikas Vepachedu and Huanyu Cheng

1 Introduction

Cyber-physical system (CPS) is an emerging generation of physical systems whose operations are governed by an integrated computational core. The ability of such real-time systems to enable interactions between its physical and computational elements is fundamental in shaping how we interact with our surrounding physical environment. The design and development of CPS play a vital role in domains varying from agriculture, energy and defense to manufacturing, aerospace and healthcare [1, 2].

In recent years, the impact of CPS in the field of manufacturing along with an exponential growth in computing capacity has given rise to what is commonly termed as *Industrie 4.0* (in Germany) or the fourth industrial revolution, which is also associated with the term Industrial Internet [3]. This revolution has the potential to enhance the efficiency of the production process, improve its productivity, and produce higher-quality products at reduced costs. The underlying concept of this incipient reform is the shift from centralized production techniques encountered in conventional manufacturing processes to a decentralized production system, enabled through CPS. This decentralization gives rise to the concept of a smart industry, which combines intelligent object networking with independent process management, and hence transforms the very nature of production [4].

The past few decades have witnessed the steady transition from traditional manufacturing techniques involving labor-intensive mechanical processes, to

---

Varun Ravikumar and Ning Yi—These authors contributed equally to this work.

V. Ravikumar · N. Yi · V. Vepachedu · H. Cheng (✉)
Department of Engineering Science and Mechanics, Materials Research Institute,
The Pennsylvania State University, University Park, PA 16802, USA
e-mail: huanyu.cheng@psu.edu

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series
in Wireless Technology, DOI 10.1007/978-3-319-42559-7_28

advanced manufacturing methods involving information technology. This is a highly integrated domain that encompasses other emerging fields such as sustainable manufacturing, supply-chain management, modeling and simulation, and business enterprising, to name a few. These broad areas when combined with CPS, enhance the concept of a smart factory, which utilizes advanced applications to optimize quality control and production efficiency [5].

In light of the recent progression of advanced manufacturing techniques, one emerging area is the domain of additive manufacturing [6], a term attributed to a variety of processes ranging from rapid prototyping and additive layer manufacturing to solid free-form fabrication and digital manufacturing. Additive manufacturing in general provides a time-efficient and economic method to manufacture complex geometries with advanced properties by minimizing material wastage. The versatility of additive manufacturing is apparent in its application across a diverse array of industries including electronics, energy, nanotechnology, aerospace and many others [6, 7].

Another innovative advanced manufacturing technique that is rapidly gaining prominence across the electronics industry is the concept of transfer printing, which complements the capability of additive manufacturing. Transfer printing is an emerging process, which primarily integrates heterogeneous materials on micro- and nano-scales for applications ranging from biomedicine [8, 9] and robotics [10, 11] to stretchable electronics [12–14] and structural health monitoring [15]. When integrated with cyber-manufacturing systems, these biomedical related applications have an underlying potential to revolutionize the healthcare industry, thus leading to improved healthcare across the world. In this chapter, we briefly review the concepts associated with the process of transfer printing in Sect. 2. The opportunities and challenges of transfer printing relevant to cyber-manufacturing systems are discussed in Sects. 3 and 4, respectively. Section 5 focuses on the future development of transfer printing in cyber-manufacturing system.

2 Fundamentals of Transfer Printing

In comparison to soft and curved biological tissues, conventional electronic devices are rigid and planar, indicating a fundamental mismatch between human tissues and electronics. In addition, it is challenging to deploy conventional electronic devices on a curvilinear surface. Stretchable electronic devices present unique solutions to address these issues by forming intimate contact between electronics and curvilinear surfaces. Capable of deforming with natural motions of the skin, this type of emerging devices can continuously monitor vital signals from the human body. Dissolvable devices that can physically disappear upon user-defined signals after functional operations, create opportunities ranging from biomedical implants to environmentally benign sensors. The ability of dissolvable electronics to completely dissolve after functioning eliminates the need for recollection of biomedical implants.

However, conventional manufacturing techniques are not compatible with manufacturing processes of stretchable and dissolvable devices. For example, stretchable sensors and actuators connected by wavy interconnects need to be assembled on soft polymeric substrates rather than rigid wafers. Dissolvable devices also have to be deployed on biodegradable substrates. The soft or biodegradable substrates are typically not compatible with conventional fabrication processes, which involve solvents and high temperature. Though additive manufacturing, such as 3D printing, presents some utility in the fabrication process of emerging sensors/actuators, the type and form of materials that can be printed are limited. Therefore, transfer printing has been introduced for fabrication of stretchable electronics and dissolvable electronics.

2.1 Basic Concepts in Transfer Printing

Transfer printing (TP) in essence represents a set of methods for the deterministic assembly of micro- and nano-materials into two- and three-dimensional, functional layouts, which are spatially organized [16]. As shown in Fig. 1, TP is broadly classified into three schemes: (a) additive transfer, (b) subtractive transfer, and (c) deterministic assembly. The third scheme is particularly powerful as it enables assembly of a broad range of materials (e.g., semiconductors, metals, and insulators) in various forms (e.g., particles, wires, ribbons, and films). TP involves the transfer of pre-fabricated micro- and nano-scale structures termed as “ink” from fabricated or growth “donor” substrates, to functional “receiver” substrates using a polymeric “stamp”. The stamps are usually made from polydimethylsiloxane (PDMS), with molded posts for selectively engaging the ink.

These inks, having a wide range of geometries and configurations, are typically constructed through conventional lithographic or other processes. Printing takes place when the “inked” stamp is brought in contact with the receiver substrate where the stamp is eventually delivered. The interaction to the stamp and substrates determines the ease from which inks are released, from the donor substrate and subsequently the stamp [17, 18].

The control of adhesion through analysis in fracture mechanics at the different interfaces involved in TP, is an important consideration that plays a fundamental role in determining an effective transfer process. In conventional TP, a key objective is to ensure strong adhesion between the stamp and ink during the “retrieval” process from a donor substrate, and weak adhesion when the ink is released onto the receiver substrate during the “printing” process (Fig. 2) [17]. Yields of transfer printing depend significantly on the ability to switch between strong and weak adhesion states in a quick and robust manner to ensure maximum efficiency. Adhesion between the stamp and the ink is largely mediated by the rate-dependent nature of the elastomer used for the stamp. The separation of stamp and ink at the interfaces is primarily understood through the initiation and propagation of cracks, initially existed at the interfaces. Griffith theory from fracture

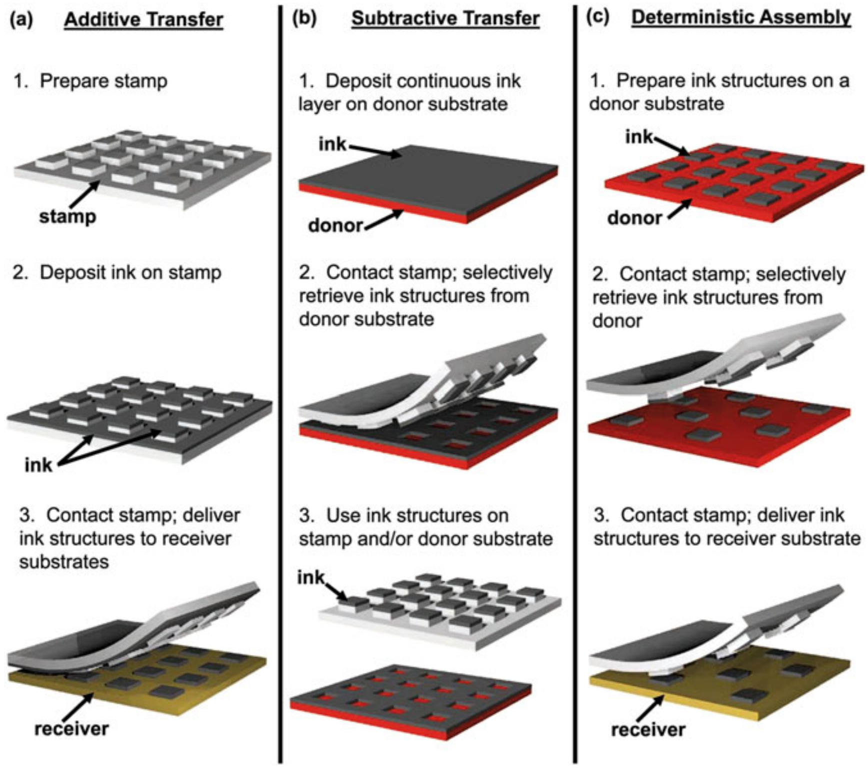


Fig. 1 Schematic of three transfer printing schemes. **a** Additive transfer delivers 'inked' materials at regions that come into contact with the target substrate. **b** Subtractive transfer results in patterned material on donor substrate, as well as inked material on stamp for additive transfer, as in (**a**). **c** Deterministic assembly involves removal of selected structures from the donor substrate, followed by a printing onto a receiving substrate. Reprinted with permission from [16]

mechanics [19] predicts that interfacial cracks initiate and propagate, when their energy release rates G exceed a critical value G_c. In general, energy release rate G of interfacial cracks is mediated by the Young's moduli of two materials at contact, contact geometry, and applied force.

For a typical TP process, adhesion between the donor substrate and ink is weaker than that between the ink and receiver substrate with an intermediate rate of adhesion between the stamp and ink. In most cases, thin adhesive layers of the interfaces involved ensure efficient transfer [20]. However, the use of viscoelastic stamps, such as PDMS, which exploits the rate-dependence of the adhesion energy at the stamp-ink interface, helps modulate adhesion through the retraction velocity. Strong van der Waals interaction between the ink and PDMS stamp mediates adhesion between the interfaces. Studies have led to the conclusion that the adhesive strength is directly proportional to the separation velocity (v) of viscoelastic stamps from a given surface [17, 21, 22]. As the energy dissipation in

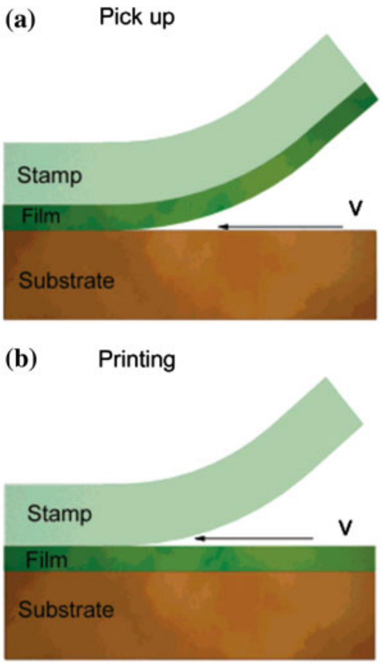


Fig. 2 Diagrams for pickup of a thin film from substrate and printing a thin film onto another substrate. Reprinted with permission from [17]

viscoelastic materials increases with increasing separation velocity, a high separation velocity yields stronger adhesion [17]. To exploit this concept, the inks are retrieved at a rapid rate from the donor substrate and then printed at a slower rate onto the receiver substrate [22, 23]. This kinetically controlled method [17, 18, 24] ensures successful transfer printing by having strong ink/stamp adhesion during pickup and weak ink/stamp adhesion during printing.

2.2 Advanced Transfer Printing Techniques

Though the separation velocity is an effective means to tune the interfacial adhesion strength, the control is limited by the range of separation velocity associated with the equipment. In order to enhance the efficacy of the TP process, one key task is to identify methods that would facilitate an increase in the modulation of adhesion at surfaces involved. Approaches developed to further enhance this aspect include laser-driven transfer printing and pneumatic-driven transfer printing. The former method makes use of a laser pulse that initiates separation at the adhesive interface due to local heating, thus causing a mismatch in thermal expansion coefficients of stamp and ink [25–27]. The latter approach modulates the adhesion involved in TP through pressurized micro-channels, near the stamp, to release inks [28]. The use of a programmable elastomeric surface allows for active control of adhesion through inflation and deflation, analogous to a balloon, of the localized regions of the stamp to maximize contact area during pickup and minimize contact area during printing.

Other widely used transfer printing techniques are inspired by modes of adhesion observed in insects and small animals such as geckos, a type of lizard with the ability to adhere to different kinds of surfaces and to rapidly change adhesion strength when required. Two mechanisms are believed to be responsible for the modulation of adhesion in geckos' feet.

First, a change in the strength of adhesion is achieved by varying contact areas. The concept is utilized in the context of TP, by distinctively designing stamps with appropriate surface relief structures, such as sharp microtips, to enable large and small contact areas during retrieval and printing, respectively [29]. The design of the stamp involves pyramidal microtips on the surface of the stamp post placed in a 2 by 2 square array, which allows the region between microtips to collapse when subjected to an external applied force (Fig. 3). During the retrieval step, the collapse of the area between the pyramidal microtips due to an applied mechanical force maximizes contact area and hence increases adhesion between the ink and stamp. The viscoelastic property of the elastomer is exploited in a manner that the elastic restoring force then returns the pyramidal microtips back to their original configuration after retraction, minimizing the adhesion strength for easy release.

The other mechanism is *directional shearing* at an interface to mechanically initiate separation, thereby controlling the adhesive behavior of components involved in this process. In this approach, the initiation of cracks at the contact areas is controlled by shear loading, where a high loading corresponds to weaker adhesion and vice versa [20, 30]. Studies involving the retraction of a single-post PDMS stamp from a silicon substrate demonstrate that the separation force decreases

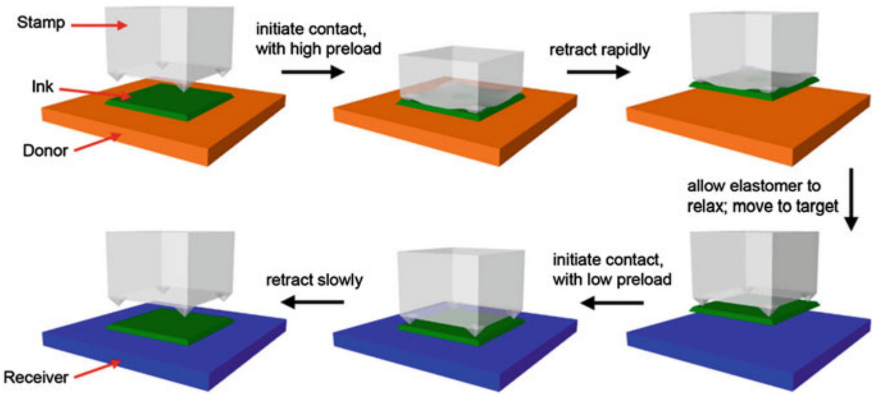


Fig. 3 Significant modulation in adhesion by employing stamps with pyramidal microtips for deterministic assembly. Reprinted with permission from [29]

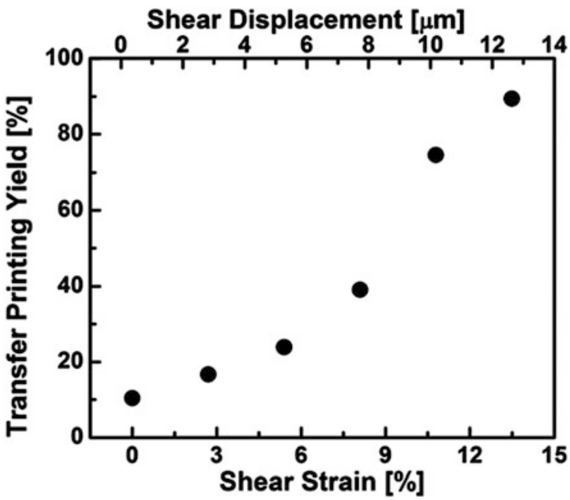


Fig. 4 Transfer yields of silicon plates onto a bare substrate of silicon wafer as a function of shear. Reprinted with permission from [20]

linearly with an increase in shear displacement. In addition, transfer yield significantly increases with an increase in shear displacement. With an increase in shear strain up to 14 %, a ten-fold increase is observed in yield (Fig. 4). Hence, shear-enhanced transfer printing, in comparison with conventional TP, has a higher efficiency and therefore facilitates the printing of inks onto complex receiver substrates.

3 Opportunities of Transfer Printing for Cyber-Manufacturing Systems

In the previous section, we summarized the basics of TP process and the advanced methods by which TP process can be enhanced for a higher yield. The process itself, has demonstrated various applications in integrated devices such as transistors [31], light emitting diodes and solar cells [32–34], flexible pressure sensor arrays [33], negative index metamaterials [35], microconcentrator photovoltaics [36], and near-field communication devices [37], through transfer printing of materials [16] including inorganic semiconductors, metals, carbons, organic materials, colloids, and biological materials. In particular, the concept of TP has gained prominence in applications pertaining to the biomedical domain, where devices have been fabricated for diagnostics/therapeutics [38] and human-machine interfaces [8, 39]. In the following section, we will discuss certain stretchable and dissolvable devices as well as the role of TP for fabricating these devices in detail.

3.1 Stretchable Electronics

Most classes of electronics used in high-end applications utilize inorganic materials such as silicon or gallium nitride in forms that are rigid and planar. When compared with the human body, which in contrast is soft and curvilinear, a mismatch in properties thus hinders the development of potential devices useful in direct correlation with the human body. A possible solution to this conundrum is the integration of organic electronic materials that are intrinsically stretchable [40–42]. Such applications are limited to devices that only require simple functions, however, as their electron mobility is not comparable with their inorganic counterparts [43]. In contrast, inorganic materials are still the best candidates for high-performance devices, with promise for commercialization as they are compatible with existing fabrication facilities for commercial devices.

Semiconductor nanomaterials used in flexible devices are usually fabricated in the form of nanomembranes, exhibiting low bending stiffness and low susceptibility to bending deformation owing to their small thicknesses [44]. Using a combination of TP techniques discussed in the previous section, the nanomembrane inks are selectively printed onto target receiver substrates using a PDMS stamp. Though inorganic nanomembranes are capable of bending to small radii of curvature, stretching of devices that are associated with the natural motion on the human body may lead to fracture due to high strain [45].

An effective strategy to accommodate these high strains is through non-linear buckling processes, where a nanomembrane is transfer printed onto a pre-stretched elastomeric substrate and the release of the pre-stretch results in the formation of a sinusoidal or herringbone pattern in 1D and 2D, respectively [46–48]. The substrate avoids fracture when stretched by amounts less than pre-strain values, which is typically ~20 %. To exploit the intrinsically fragile inorganic materials such as silicon (fracture strain ~1 %) for even larger stretchability, structural design takes the concept of assembly electronic components connected by wavy interconnects on a soft polymeric substrate. In such design, electronic components are fully bonded to the substrate, whereas the wavy interconnects are freestanding [14, 49, 50]. Upon stretching, the wavy interconnects move out-of-plane to accommodate the large deformation applied to the system [51, 52]. The device components as well as the portion of substrate bonded beneath experience negligible deformation, preventing damage to these delicate components.

Recently developed stretchable devices have demonstrated various applications in the biomedical domain ranging from drug delivery [53, 54] and energy harvesting [55–57] to wearable sensors that can sense temperature [58–60], strain [8, 61, 62], skin hydration level [63–65], and electrophysiological signals [8, 10, 39, 52, 66]. The continuous monitoring capability enabled by this type of emerging devices will have a profound role in transforming the healthcare industry, with a shift of focus from hospital-centered to person-centered.

In addition, transfer printing plays a critical role to deliver electronic devices onto any curvilinear surface of interest [67, 68]. This capability goes beyond the

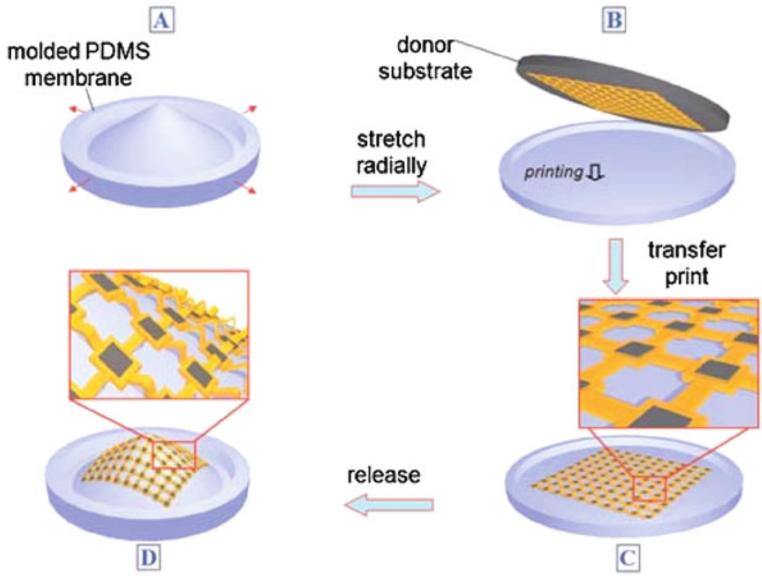


Fig. 5 Deployment of electronic systems onto an arbitrary curvilinear surface via transfer printing. Reprinted with permission from [68]

applications in bio-integrated devices that require intimate contact to tissues of complex topologies [69]. It can also be applied to sensors for structural health monitoring, particularly at regions with curvilinear surfaces [70]. The process starts with casting a mold (Fig. 5a) on the target surface, followed by radial tension to stretch the mold into a flat shape (Fig. 5b) on which electronic systems can be transfer printed (Fig. 5c). With electronic components bonded to the mold and free for interconnects between adjacent components, release of the pre-stretch transform the electronic systems into the target shape (Fig. 5d), which can then be integrated with ease.

3.2 *Dissolvable Electronics*

Dissolvable (or transient) electronics are an emerging class of electronics whose defining characteristic is the ability to dissolve in a predetermined time and manner. These electronics can have applications in a variety of domains. They can be used for biomedical implants, for instance, in which case the need for a surgery to remove the device would be unnecessary [9, 71, 72], as it would simply dissolve on its own into harmless end products (Fig. 6a). They can also provide another step towards eco-friendly electronics by designing biodegradable devices; upon contact

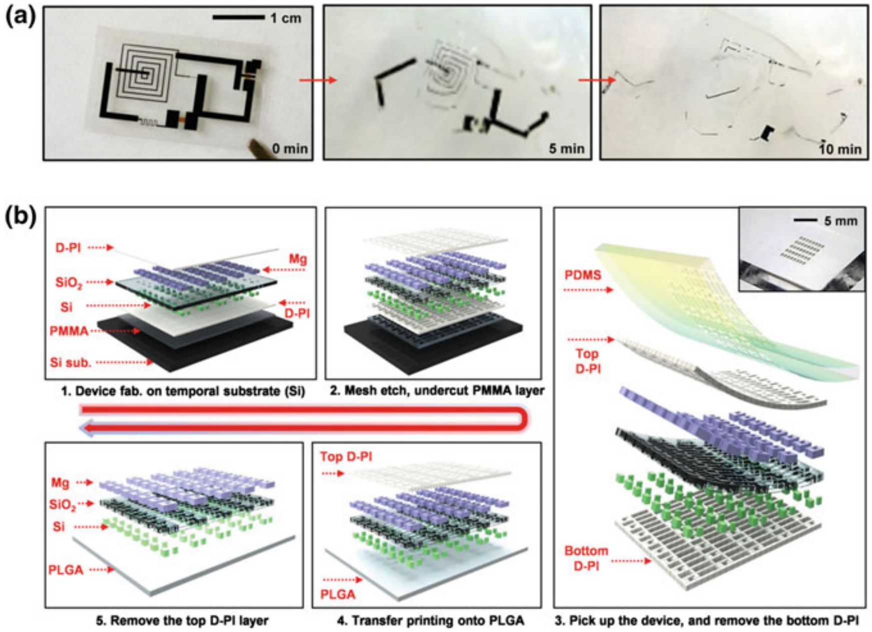


Fig. 6 **a** Optical images show the dissolution sequence of a proof-of-concept device in deionized water. Reprinted with permission from [81]. **b** Materials and procedures to fabricate dissolvable electronic devices on biodegradable substrates. Reprinted with permission from [85]

with water, the said devices would disintegrate into harmless end products and be dispersed into the environment.

Numerous factors such as pH [73], temperature [73], porosity [74, 75], ion concentration [76], and many others [77, 78] have to be considered when constructing such devices. Through consideration of these factors on both a qualitative and mathematical level, transient electronic devices can be built to dissolve in a predictable manner in accordance with certain essential constraints. Ongoing research is also in progress to unearth novel ways to actively control [79, 80] the dissolution of the device as opposed to allowing it to dissolve by itself.

Some of the common materials used in dissolvable electronics include inorganic semiconductors (e.g., silicon [81, 82], germanium and an alloy of silicon and germanium [74], zinc oxides [55]), transient metals (e.g., magnesium, zinc, iron, tungsten, molybdenum) [83], and insulators (e.g., magnesium oxides [81], silk [84], silicon oxides and nitrides [75]). These materials are often used in the form of thin films to minimize the amount of materials to dissolve, dictating rapid dissolution.

As with almost all electronics, the circuitry requires a substrate on which the components and connections can be placed in an orderly fashion. For conventional electronics this substrate is often rigid such as silicon wafers. However, for dissolvable electronics, at least in their biomedical applications, while a wafer is initially used to create the circuitry, the final product uses a very thin, flexible substrate composed of a water-soluble material such as silk [81]. In order to transfer a layer of circuitry onto such a substrate, TP can be used quite reliably [85]. One method of performing this transfer printing is as follows (Fig. 6b) [85]. First, wafer coated with a sacrificial layer is used as a temporary substrate on which the device is fabricated. Etching of the sacrificial layer releases the device from the temporary substrate and this device is then picked up onto a stamp. The stamp is then used to transfer print the circuitry onto the biodegradable substrate such as PLGA. Since the thin biodegradable substrate can be delicate, the methods that exert too much stress or strain on the film are not desirable. On the other hand, TP allows for the physical application of circuitry onto such a delicate substrate.

In some other cases, the deployment of an actual transient electronic device may use the idea of TP by printing the circuitry onto the target locations. For instance, in one particular study, a dissolvable electronic device consisting of an electrode array and a silk substrate was placed onto the brain in a way that resembles TP. The electrode array was placed on the brain with the silk substrate after which the silk substrate was dissolved using a saline solution, thereby transferring a circuit onto the brain [86]. While it deviates from conventional TP, this process yields an identical result to that observed from conventional TP. In this case, not only is TP useful for transferring the device to the subject but, to an extent, the issue of adhesion is also largely addressed with fewer extra measures to keep the device in place.

In the future, it can be expected that the TP technique will continue to be useful in the fabrication of dissolvable electronics. It is useful due to its ease of use relative to other methods and its ability to gently transfer a pattern without placing too much stress on the material in question. Once again, this gentle quality will prove useful in the context of medicine where the device must be placed onto a living organism without extensive surgery or the risk of cell damage. In addition, TP is bound to be significant in the context of environmentally benign devices [87]. The rapid advancement in technologies results in obsolescence of devices. The created electronic wastes pose significant environmental concerns. With the capability to dissolve without trace, environmentally benign devices could potentially reduce the electronic wastes, thereby providing a solution to the grand challenge.

3.3 Opportunities of Transfer Printing Enabled Devices for Cyber-Manufacturing Systems

Capable of conforming to arbitrary curvilinear surfaces, stretchable sensors can be integrated into manufacturing equipment with complex geometries. The data streamed from these sensors would provide critical information for health conditions of the manufacturing equipment, thereby enabling necessary feedbacks in the manufacturing processes. In addition, advancement in technology creates rapid obsolescence for electronic devices, which poses significant environmental concerns. The introduction of dissolvable electronics can reduce electronic wastes, thereby providing a viable means for sustainable manufacturing. Hence, introducing transfer printing to cyber-manufacturing systems creates new opportunities for improved systems with sustainable designs.

However, scalable manufacturing with transfer printing still remains a distant prospect, which hinders the large-scale manufacturing of stretchable devices and dissolvable electronics. Cyber-manufacturing systems, featuring the interaction between computation and physical control as well as enabled communication between human and machines, make it suitable for large-scale manufacturing. Combining transfer printing and cyber-manufacturing systems provides a viable pathway for the future development of stretchable and dissolvable electronics in areas of healthcare, industrial equipment monitoring, and environmentally benign electronics.

4 Challenges of Transfer Printing for Cyber-Manufacturing Systems

It is challenging to change from the conventional manual process to an automatic machine-controlled manner for transfer printing. Some problems that need to be considered include the design of the fabrication process, the modulation of adhesion strength, and the inherent challenges of the cyber-physical interactions.

As discussed in advanced TP techniques, the ability to modulate the degree of adhesion between inks and the stamp is of prime importance in moderating the efficiency of a TP process. Bio-inspiration provides relatively simple yet powerful design principles, as evidenced in the advanced TP techniques, but the enabled capabilities are still far from their counterparts in nature. For example, it is challenging to combine the shear-enhanced strategy [20] and surface relief structures [29, 88, 89] to mimic the versatile modulation of adhesion observed in geckos.

In addition, it is desirable to integrate TP with large-scale manufacturing for cyber-manufacturing systems. To address this goal, angled posts are introduced in roller modes of operation for TP [89]. Studies inspecting the feasibility of such a system demonstrate the ability to switch adhesion modes depending on the direction of rolling. The use of angled posts on the stamp surface regulates the tendency towards crack formation and hence controls adhesion. Adhesion strength for such designs can be modulated by the direction of retraction due to the presence of asymmetric contact angles with a flat substrate. The ability to switch adhesion states when using angled posts enables efficient TP onto certain complex substrates.

When angled posts are placed on the surface of rollers (Fig. 7a), the control in rolling directions replaces the control in directions of retraction for integration on planar surfaces. In the case of forward rolling in a clockwise direction, the angled posts form a small contact angle with the substrate, whereas backward rolling in a counter-clockwise direction involves larger contact (Fig. 7b). Experiments indicate that for a constant rolling speed, the energy release rate that characterizes the

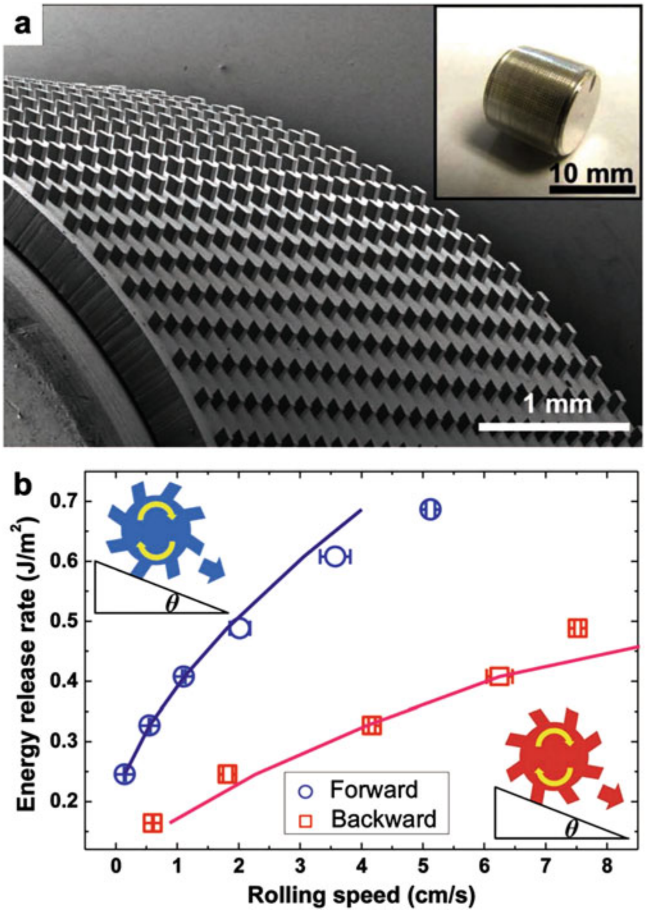


Fig. 7 Demonstration of rotation direction-dependent adhesion with angled posts integrated on a roller, with promising applications in roll-to-roll manufacturing. **a** SEM image and optical image (*inset*) of a stamp roller coated with an array of angled posts. **b** Experimental and theoretical results in energy release rate of an angled post roller as a function of terminal rolling speed for forward (high adhesion) and backward (low adhesion) rolling, respectively. Reprinted with permission from [89]

condition of interface separation for forward rolling is greater than that for backward rolling, thus signifying the suitability of forward rolling for a retrieval step and backward rolling for a printing step.

To achieve high throughput rates in TP, instead of transfer printing a single device at a time, we can utilize the roll-to-roll application enabled by integrating angled posts on rollers for cost effective mass production at an enhanced rate. The inherent parallelism in a TP process allows for the printing of desired number of devices (inks), which are usually on the micro- to nano-scales, onto the receiving substrate with an order on the centimeter-scale, in a single print cycle. However, increasing parallelism to satisfy the demands of the manufacturing market can be challenging to monitor and regulate, owing to the possibility of errors in the form of angular misalignment, stamp collapse, receiving substrate defects and microstructure residuals, all of which can result in local printing failures. Hence, there is a need to identify techniques to ensure an effective and robust system to monitor and minimize the possibility of errors during the TP process. One possible approach to solve this conundrum is the integration of a computational system with the transfer printing process to efficiently monitor and hence enhance the efficacy of the system.

To integrate the TP process with cyber-physical systems, we need to consider the intrinsic challenges of the cyber-physical systems. Combining the computation and physical process with feedback loops to affect each other [90], manufacturing based on cyber-physical systems is expected to be versatile, real-time control, robust, autonomy, self-maintained, reconfigurable, transparent, predictable, and global networked [91]. Designing a cyber-manufacturing system with such expectations requires the engineers to solve several challenges, including the robustness and reliability of a complex system, real-time interaction between the virtual and real worlds, validation and certification of the manufacture systems, and integration of the system by components provided by different suppliers. Representative challenges of cyber-manufacturing systems are addressed below.

Challenge 1: Robustness and Reliability

A cyber-physical system will not be operating in a totally controlled environment without any disturbance. The designed cyber-physical system should be robust to unexpected events and subsystem failures, which has been a long-lasting problem for engineers. It is easy to construct a reliable and predictable system with reliable and predictable components. But the components in the real world are subject to unexpected disturbances. For cyber-manufacturing systems designed for transfer printing, it will also be challenging to resolve the unexpected circumstances during the fabrication process. For example, defects of the ink fabricated on the donor substrate may result in subsequent failures in the designed device. A robust system is required to detect this unsuccessful process and predict the failure in fabrication processes. The robustness of cyber-manufacturing systems is of vital importance for the successful implementation of it in industry.

Challenge 2: Real-Time Control

The computation based on software needs to be aligned with the time lapse in the physical world. The computation is expected to provide real-time feedback for effective interaction between the computation and physical process. However, no such programming language and concurrent software are available for this purpose [90]. When designing a cyber-manufacturing system for transfer printing, it is important to ensure real-time control for each manufacturing step. With networked and versatile features for a cyber-manufacturing system, this challenge can be even more difficult to address.

Challenge 3: High Cost in Verification and Validation
Engineers have always been challenged to verify that the constructed systems can perform properly through different methods of verification and validation [92]. Although it can provide evidence for the reliability and feasibility of a newly designed system, the time and resources required for such process can be prohibitively high, especially when the system is complicated [92]. For example, verification and validation of a new safety-critical system in the aviation industry would consume more than 50 % of the resources required to develop such a system [93]. Therefore, it is critical to strike a balance between the cost and versatility of a cyber-manufacturing system for transfer printing.

Challenge 4: Integration with Various Subsystems
New systems are often constructed by integrating various subsystems. As a typical example, a vehicle control system is composed of various hardware and software subsystems supplied by different vendors [93]. The integration of subsystems designed by various manufacturers often requires a redesign of the system interface. This remains an issue for cyber-manufacturing systems, particularly the one for transfer printing with high expectations in robustness, real-time control, and multi-functionality.

5 Future Scope

Enabling a cost-effective transfer printing process through a cyber-manufacturing system promises great potential for biomedical devices. As demonstrated in Sect. 3, the emerging concept of transfer printing can be widely used in a variety of fields such as electronics and optoelectronics, for suitable end applications in clinical research and consumer domains. The immense potential of devices fabricated from TP techniques in the biomedical field, due to its ability to seamlessly integrate with soft human tissue, can revolutionize the very future of the healthcare industry. Though the current research involving the designing and realization of this concept is in its neonate stage, there exists limitless opportunities to enhance and thus improve the efficiency and affordability of such devices in the near future through the development of cyber-manufacturing systems. Although we observe an emerging growth in the market of healthcare devices such as smart-watches and headbands, their relevance to improving human health will indubitably spark an immense global market to offer affordable healthcare devices in the near future. With the rising cost of healthcare worldwide, there is a growing need for affordable but quality healthcare services, and the necessity to identify reliable medical device cyber-physical architectures to address this problem is a major cause for concern.

For the future development of such a cyber-manufacturing system for transfer printing, researchers and engineers should innovate on both the transfer printing process and the cyber-physical systems design. Innovative methods in materials and device engineering along with novel ideas in structural designs that can mimic certain phenomena in nature represent some of the imminent objectives to be addressed in this field. However, a few of the immediate challenges involved during the transfer printing of required devices, include modifications to the size and shape of the device, processing parameters set-up for varies materials and process, and the validation of the feasibility of such automotive process should also be addressed. Current studies utilize TP techniques to design and fabricate planar devices on the micro-scale range. However, if we were to further reduce the size to nano-scale in a controllable manner, possibilities of enhanced efficiency through an increase in computational power and increased manufacturing output exist, which thus would regulate the cost of such devices. Moreover, so far we have been only able to construct devices that are essentially planar in nature. A practical yet interesting target would be to utilize existing or new TP techniques to manufacture devices in three dimensional geometries [94], which then could potentially be used for a wide variety of novel applications. The incorporation of computational systems to automate and regulate the manufacturing process would also immensely augment the efficiency of the entire system, as a means to address a major challenge that is to design a robust system with minimal errors to actively monitor and control critical aspects of the processes involved. Issues such as the temporal problem of the virtual part, the testing of such a cyber-physical system, optimization of interfaces between different subsystems should also be an objective for both the academia and industry.

Acknowledgments The authors acknowledge the start-up fund provided by the Engineering Science and Mechanics Department, College of Engineering, and Materials Research Institute at The Pennsylvania State University.

References

1. Rajkumar RR, Lee I, Sha L, Stankovic J (2010) Cyber-physical systems: the next computing revolution. In: Proceedings of the 47th design automation conference, pp 731–736
2. Wolf M (2014) High-performance embedded computing: applications in cyber-physical systems and mobile computing. Newnes
3. Lohr S (2012) The age of big data. New York Times, vol 11
4. Kagermann H, Helbig J, Hellinger A, Wahlster W (2013) Recommendations for implementing the strategic initiative INDUSTRIE 4.0: securing the future of german manufacturing industry; final report of the Industrie 4.0 working group, Forschungsunion
5. Shipp SS, Gupta N, Lal B, Scott JA, Weber CL, Finnin MS et al (2012) Emerging global trends in advanced manufacturing, DTIC Document
6. Gibson I, Rosen DW, Stucker B (2010) Additive manufacturing technologies. Springer, US
7. Huang Y, Leu MC (2014) Frontiers of additive manufacturing research and education
8. Kim D-H, Lu N, Ma R, Kim Y-S, Kim R-H, Wang S et al (2011) Epidermal electronics. Science 333:838–843
9. Kang S-K, Murphy RK, Hwang S-W, Lee SM, Harburg DV, Krueger NA et al (2016) Bioresorbable silicon electronic sensors for the brain. Nature 530:71–76
10. Jeong JW, Yeo WH, Akhtar A, Norton JJ, Kwack YJ, Li S et al (2013) Materials and optimized designs for human-machine interfaces via epidermal electronics. Adv Mater 25:6839–6846

11. Ying M, Bonifas AP, Lu N, Su Y, Li R, Cheng H et al (2012) Silicon nanomembranes for fingertip electronics. Nanotechnology 23:344004
12. Rogers JA, Someya T, Huang YG (2010) Materials and mechanics for stretchable electronics. Science 327:1603–1607
13. Kim DH, Song JZ, Choi WM, Kim HS, Kim RH, Liu ZJ et al (2008) Materials and noncoplanar mesh designs for integrated circuits with linear elastic responses to extreme mechanical deformations. Proc Natl Acad Sci USA 105:18675–18680
14. Kim DH, Ahn JH, Choi WM, Kim HS, Kim TH, Song JZ et al (2008) Stretchable and foldable silicon integrated circuits. Science 320:507–511
15. Kang I, Schulz MJ, Kim JH, Shanov V, Shi D (2006) A carbon nanotube strain sensor for structural health monitoring. Smart Mater Struct 15:737
16. Carlson A, Bowen AM, Huang Y, Nuzzo RG, Rogers JA (2012) Transfer printing techniques for materials assembly and micro/nanodevice fabrication. Adv Mater 24:5284–5318
17. Feng X, Meitl MA, Bowen AM, Huang Y, Nuzzo RG, Rogers JA (2007) Competing fracture in kinetically controlled transfer printing. Langmuir 23:12555–12560
18. Meitl MA, Zhu ZT, Kumar V, Lee KJ, Feng X, Huang YY et al (2006) Transfer printing by kinetic control of adhesion to an elastomeric stamp. Nat Mater 5:33–38
19. Anderson TL, Anderson T (2005) Fracture mechanics: fundamentals and applications. CRC press, Boca Raton
20. Carlson A, Kim-Lee H-J, Wu J, Elvikis P, Cheng H, Kovalsky A et al (2011) Shear-enhanced adhesiveless transfer printing for use in deterministic materials assembly. Appl Phys Lett 98:264104
21. Feng X, Cheng H, Bowen AM, Carlson AW, Nuzzo RG, Rogers JA (2013) A finite-deformation mechanics theory for kinetically controlled transfer printing. J Appl Mech 80:061023
22. Cheng H, Li M, Wu J, Carlson A, Kim S, Huang Y et al (2013) A viscoelastic model for the rate effect in transfer printing. J Appl Mech 80:041019
23. Chen H, Feng X, Huang Y, Huang Y, Rogers JA (2013) Experiments and viscoelastic analysis of peel test with patterned strips for applications to transfer printing. J Mech Phys Solids 61:1737–1752
24. Kim TH, Carlson A, Ahn JH, Won SM, Wang SD, Huang YG et al (2009) Kinetically controlled, adhesiveless transfer printing using microstructured stamps. Appl Phys Lett 94:113502
25. Li R, Li Y, Lü C, Song J, Saeidpourazar R, Fang B et al (2012) Thermo-mechanical modeling of laser-driven non-contact transfer printing: two-dimensional analysis. Soft Matter 8:3122–3127
26. Li R, Li Y, Lü C, Song J, Saeidpourazar R, Fang B et al (2012) Axisymmetric thermo-mechanical analysis of laser-driven non-contact transfer printing. Int J Fract 176:189–194
27. Saeidpourazar R, Sangid MD, Rogers JA, Ferreira PM (2012) A prototype printer for laser driven micro-transfer printing. J Manuf Process 14:416–424
28. Carlson A, Wang S, Elvikis P, Ferreira PM, Huang Y, Rogers JA (2012) Active, programmable elastomeric surfaces with tunable adhesion for deterministic assembly by transfer printing. Adv Funct Mater 22:4476–4484
29. Kim S, Wu JA, Carlson A, Jin SH, Kovalsky A, Glass P et al (2010) Microstructured elastomeric surfaces with reversible adhesion and examples of their use in deterministic assembly by transfer printing. Proc Natl Acad Sci USA 107:17095–17100
30. Cheng H, Wu J, Yu Q, Kim-Lee H-J, Carlson A, Turner KT et al (2012) An analytical model for shear-enhanced adhesiveless transfer printing. Mech Res Commun 43:46–49
31. Liao L, Bai J, Qu Y, Lin Y-C, Li Y, Huang Y et al (2010) High-κ oxide nanoribbons as gate dielectrics for high mobility top-gated graphene transistors. Proc Natl Acad Sci 107: 6711–6715
32. Kim T-I, Lee SH, Li Y, Shi Y, Shin G, Lee SD et al (2014) Temperature-and size-dependent characteristics in ultrathin inorganic light-emitting diodes assembled by transfer printing. Appl Phys Lett 104:051901

33. Yoon J, Baca AJ, Park SI, Elvikis P, Geddes JB, Li LF et al (2008) Ultrathin silicon solar microcells for semitransparent, mechanically flexible and microconcentrator module designs. Nat Mater 7:907–915
34. Kim RH, Bae MH, Kim DG, Cheng H, Kim BH, Kim DH et al (2011) Stretchable, transparent graphene interconnects for arrays of microscale inorganic light emitting diodes on rubber substrates. Nano Lett 11:3381–3886
35. Chanda D, Shigeta K, Gupta S, Cain T, Carlson A, Mihi A et al (2011) Large-area flexible 3D optical negative index metamaterial formed by nanotransfer printing. Nat Nanotechnol 6:402–407
36. Furman B, Menard E, Gray A, Meitl M, Bonafede S, Kneeburg D et al (2010) A high concentration photovoltaic module utilizing micro-transfer printing and surface mount technology. In: 2010 35th IEEE Photovoltaic Specialists Conference (PVSC), pp 000475–000480
37. Kim J, Banks A, Cheng H, Xie Z, Xu S, Jang KI et al (2015) Epidermal electronics with advanced capabilities in near—field communication. small 11:906–912
38. Hattori Y, Falgout L, Lee W, Jung SY, Poon E, Lee JW et al (2014) Multifunctional skin-like electronics for quantitative, clinical monitoring of cutaneous wound healing. Adv Healthc Mater 3:1597–1607
39. Norton JJ, Lee DS, Lee JW, Lee W, Kwon O, Won P et al (2015) Soft, curved electrode systems capable of integration on the auricle as a persistent brain–computer interface. Proc Natl Acad Sci 112:3920–3925
40. Lee TW, Zaumseil J, Bao ZN, Hsu JWP, Rogers JA (2004) Organic light-emitting diodes formed by soft contact lamination. Proc Natl Acad Sci USA 101:429–433
41. Sekitani T, Nakajima H, Maeda H, Fukushima T, Aida T, Hata K et al (2009) Stretchable active-matrix organic light-emitting diode display using printable elastic conductors. Nat Mater 8:494–499
42. Sekitani T, Noguchi Y, Hata K, Fukushima T, Aida T, Someya T (2008) A rubberlike stretchable active matrix using elastic conductors. Science 321:1468–1472
43. Hwang SW, Kim DH, Tao H, Kim Ti, Kim S, Yu KJ et al (2013) Materials and fabrication processes for transient and bioresorbable high-performance electronics. Adv Funct Mater 23:4087–4093
44. Khang DY, Jiang HQ, Huang Y, Rogers JA (2006) A stretchable form of single-crystal silicon for high-performance electronics on rubber substrates. Science 311:208–212
45. Viventi J, Kim DH, Moss JD, Kim YS, Blanco JA, Annetta N et al (2010) A conformal, bio-interfaced class of silicon electronics for mapping cardiac electrophysiology. Sci Transl Med 2:24ra22
46. Sun YG, Choi WM, Jiang HQ, Huang YGY, Rogers JA (2006) Controlled buckling of semiconductor nanoribbons for stretchable electronics. Nat Nanotechnol 1:201–207
47. Cheng H, Song J (2013) A simply analytic study of buckled thin films on compliant substrates. J Appl Mech 81:024501
48. Cheng H, Zhang Y, Hwang K-C, Rogers JA, Huang Y (2014) Buckling of a stiff thin film on a pre-strained bi-layer substrate. Int J Solids Struct 51:3113–3118
49. Liu Z, Cheng H, Wu J (2014) Mechanics of solar module on structured substrates. J Appl Mech 81:064502
50. Zhang Y, Fu H, Su Y, Xu S, Cheng H, Fan JA et al (2013) Mechanics of ultra-stretchable self-similar serpentine interconnects. Acta Mater 61:7816–7827
51. Xu S, Zhang Y, Cho J, Lee J, Huang X, Jia L et al (2013) Stretchable batteries with self-similar serpentine interconnects and integrated wireless recharging systems. Nat Commun 4:1543
52. Jeong JW, Kim MK, Cheng H, Yeo WH, Huang X, Liu Y et al (2014) Capacitive epidermal electronics for electrically safe, long-term electrophysiological measurements. Adv Healthc Mater 3:642–648
53. Honda W, Harada S, Arie T, Akita S, Takei K (2014) Wearable, human-interactive, health-monitoring, wireless devices fabricated by macroscale printing techniques. Adv Funct Mater 24:3299–3304

54. Son D, Lee J, Qiao S, Ghaffari R, Kim J, Lee JE et al (2014) Multifunctional wearable devices for diagnosis and therapy of movement disorders. Nat Nanotechnol 9:397–404
55. Dagdeviren C, Hwang SW, Su Y, Kim S, Cheng H, Gur O et al (2013) Transient, biocompatible electronics and energy harvesters based on ZnO. Small 9:3398–3404
56. Dagdeviren C, Yang BD, Su Y, Tran PL, Joe P, Anderson E et al (2014) Conformal piezoelectric energy harvesting and storage from motions of the heart, lung, and diaphragm. Proc Natl Acad Sci 111:1927–1932
57. Lee M, Chen CY, Wang S, Cha SN, Park YJ, Kim JM et al (2012) A hybrid piezoelectric structure for wearable nanogenerators. Adv Mater 24:1759–1764
58. Webb RC, Bonifas AP, Behnaz A, Zhang Y, Yu KJ, Cheng H et al (2013) Ultrathin conformal devices for precise and continuous thermal characterization of human skin. Nat Mater 12:938–944
59. Xu L, Gutbrod SR, Bonifas AP, Su Y, Sulkin MS, Lu N et al (2014) 3D multifunctional integumentary membranes for spatiotemporal cardiac measurements and stimulation across the entire epicardium. Nat Commun 5:3329
60. Fan JA, Yeo WH, Su Y, Hattori Y, Lee W, Jung SY et al (2014) Fractal design concepts for stretchable electronics. Nat Commun 5:3266
61. Yamada T, Hayamizu Y, Yamamoto Y, Yomogida Y, Izadi-Najafabadi A, Futaba DN et al (2011) A stretchable carbon nanotube strain sensor for human-motion detection. Nat Nanotechnol 6:296–301
62. Shi J, Li X, Cheng H, Liu Z, Zhao L, Yang T et al (2016) Graphene reinforced carbon nanotube networks for wearable strain sensors. Adv Funct Mater
63. Huang X, Cheng H, Chen K, Zhang Y, Zhang Y, Liu Y et al (2013) Epidermal impedance sensing sheets for precision hydration assessment and spatial mapping. IEEE Trans Biomed Eng 60:2848–2857
64. Huang X, Liu Y, Cheng H, Shin WJ, Fan JA, Liu Z et al (2014) Materials and designs for wireless epidermal sensors of hydration and strain. Adv Funct Mater 24:3846–3854
65. Cheng H, Zhang Y, Huang X, Rogers JA, Huang Y (2013) Analysis of a concentric coplanar capacitor for epidermal hydration sensing. Sens Actuators A Phys 203:149–153
66. Jang KI, Han SY, Xu S, Mathewson KE, Zhang Y, Jeong JW et al (2014) Rugged and breathable forms of stretchable electronics with adherent composite substrates for transcutaneous monitoring. Nat Commun 5:4779
67. Ko HC, Stoykovich MP, Song JZ, Malyarchuk V, Choi WM, Yu CJ et al (2008) A hemispherical electronic eye camera based on compressible silicon optoelectronics. Nature 454:748–753
68. Wang SD, Xiao JL, Song JZ, Ko HC, Hwang KC, Huang YG et al (2010) Mechanics of curvilinear electronics. Soft Matter 6:5757–5763
69. Jang K-I, Chung HU, Xu S, Lee CH, Luan H, Jeong J et al (2015) Soft network composite materials with deterministic and bio-inspired designs. Nat Commun 6
70. Ko HC, Shin G, Wang SD, Stoykovich MP, Lee JW, Kim DH et al (2009) Curvilinear electronics formed using silicon membrane circuits and elastomeric transfer elements. Small 5:2703–2709
71. Yu KJ, Kuzum D, Hwang S-W, Kim BH, Juul H, Kim NH et al (2016) Bioresorbable silicon electronics for transient spatiotemporal mapping of electrical activity from the cerebral cortex. Nat Mater
72. Hwang S-W, Lee CH, Cheng H, Jeong J-W, Kang S-K, Kim J-H et al (2015) Biodegradable elastomers and silicon nanomembranes/nanoribbons for stretchable, transient electronics, and biosensors. Nano Lett 15:2801–2808
73. Hwang SW, Park G, Cheng H, Song JK, Kang SK, Yin L et al (2014) 25th anniversary article: materials for high-performance biodegradable semiconductor devices. Adv Mater 26: 1992–2000
74. Kang S-K, Park G, Kim K, Hwang S-W, Cheng H, Shin J et al (2015) Dissolution chemistry and biocompatibility of silicon-and germanium-based semiconductors for transient electronics. ACS Appl Mater Interfaces 7:9297–9305

75. Kang S-K, Hwang S-W, Cheng H, Yu S, Kim BH, Kim J-H et al (2014) Dissolution behaviors and applications of silicon oxides and nitrides in transient electronics. Adv Funct Mater 24:4427–4434
76. Yin L, Farimani AB, Min K, Vishal N, Lam J, Lee YK et al (2015) Mechanisms for hydrolysis of silicon nanomembranes as used in bioresorbable electronics. Adv Mater 27:1857–1864
77. Fu K, Wang Z, Dai J, Carter M, Hu L (2016) Transient electronics: materials and devices. Chem Mater 28:3527–3539
78. Cheng H, Vepachedu V (2016) Recent development of transient electronics. Theor Appl Mech Lett 6:21–31
79. Park CW, Kang SK, Hernandez HL, Kaitz JA, Wie DS, Shin J et al (2015) Thermally triggered degradation of transient electronic devices. Adv Mater 27:3783–3788
80. Hernandez HL, Kang SK, Lee OP, Hwang SW, Kaitz JA, Inci B et al (2014) Triggered transience of metastable poly (phthalaldehyde) for transient electronics. Adv Mater 26: 7637–7642
81. Hwang SW, Tao H, Kim DH, Cheng H, Song JK, Rill E et al (2012) A physically transient form of silicon electronics. Science 337:1640–1644
82. Hwang SW, Park G, Edwards C, Corbin EA, Kang SK, Cheng H et al (2014) Dissolution chemistry and biocompatibility of single-crystalline silicon nanomembranes and associated materials for transient electronics. ACS Nano 8:5843–5851
83. Yin L, Cheng H, Mao S, Haasch R, Liu Y, Xie X et al (2014) Dissolvable metals for transient electronics. Adv Funct Mater 24:645–658
84. Brenckle MA, Cheng H, Hwang S, Tao H, Paquette M, Kaplan DL et al (2015) Modulated degradation of transient electronic devices through multilayer silk fibroin pockets. ACS Appl Mater Interfaces 7:19870–19875
85. Hwang SW, Song JK, Huang X, Cheng H, Kang SK, Kim BH et al (2014) High-performance biodegradable/transient electronics on biodegradable polymers. Adv Mater 26:3905–3911
86. Kim DH, Viventi J, Amsden JJ, Xiao JL, Vigeland L, Kim YS et al (2010) Dissolvable films of silk fibroin for ultrathin conformal bio-integrated electronics. Nat Mater 9:511–517
87. Jung YH, Chang TH, Zhang H, Yao C, Zheng Q, Yang VW et al (2015) High-performance green flexible electronics based on biodegradable cellulose nanofibril paper. Nat Commun 6:7170
88. Kim S, Carlson A, Cheng H, Lee S, Park J-K, Huang Y et al (2012) Enhanced adhesion with pedestal-shaped elastomeric stamps for transfer printing. Appl Phys Lett 100:171909
89. Yang SY, Carlson A, Cheng H, Yu Q, Ahmed N, Wu J et al (2012) Elastomer surfaces with directionally dependent adhesion strength and their use in transfer printing with continuous roll-to-roll applications. Adv Mater 24:2117–2122
90. Lee EA (2008) Cyber physical systems: design challenges. In: 11th IEEE symposium on object/component/service-oriented real-time distributed computing, ISORC 2008, pp 363–369
91. Monostori L (2014) Cyber-physical production systems: roots, expectations and R&D challenges. In: 47th CIRP conference on manufacturing systems, pp 9–13
92. High Confidence Software and Systems Coordinating Group (2009) High-confidence medical devices: cyber-physical systems for 21st century health care. A research and development needs report
93. Baheti R, Gill H (2010) Cyber-physical systems. In: Samad T, Annaswamy A (eds) The impact of control technology. IEEE Control Systems Society, pp 161–166
94. Xu S, Yan Z, Jang K-I, Huang W, Fu H, Kim J et al (2015) Assembly of micro/nanomaterials into complex, three-dimensional architectures by compressive buckling. Science 347:154–159

Advanced Manufacturing Innovation Ecosystems: The Case of Massachusetts

Yilmaz Uygun and Elisabeth Beck Reynolds

1 Introduction

This chapter looks at manufacturing companies' innovation capacity as it relates to cyber-physical production systems from a broader innovation ecosystem perspective. This contribution is guided by the research question which systemic measures need to be implemented to ensure a proper and wider diffusion of such complex systems by the example of Massachusetts and its advanced manufacturing companies that are poised to embrace cyber-physical production systems.

Recent years have brought a renewed focus on the importance of manufacturing to the health and future growth of the U.S. economy. Indeed, several studies and public-private initiatives have highlighted the need to maintain and build manufacturing capabilities to support economic growth, good jobs, and national security. Perhaps most importantly, they have linked the nation's manufacturing capabilities to its ability to innovate. Advanced manufacturing is essential for developing new products and processes across a range of industries, both established and emerging. As others have pointed out, the loss of these capabilities can shift an industry's center of gravity as higher value-added activities follow manufacturing abroad.

In few states is the link between manufacturing and innovation more evident than in Massachusetts. While manufacturing represents only 9 % of employment in the Commonwealth (approximately 250,000 jobs), compared to 11 % in the U.S. overall, it is integral to several of the state's most important industries, including aerospace/defense, semiconductors and computers, biopharmaceuticals, and medical devices. Massachusetts manufacturers compete globally on their innovation capacity, high skills, product quality, and rapid response.

Y. Uygun (✉) · E.B. Reynolds
Industrial Performance Center, Massachusetts Institute of Technology,
Cambridge, USA
e-mail: uygun@mit.edu

S. Jeschke et al. (eds.), *Industrial Internet of Things*, Springer Series
in Wireless Technology, DOI 10.1007/978-3-319-42559-7_29

Small and medium-sized enterprises (SMEs)[1] play a critical role in maintaining and growing the manufacturing strengths of the U.S. and Massachusetts economies. These companies are the “backbone” of the country’s and the region’s industrial capabilities and they exist in every community where manufacturing takes place. SMEs supply both the large established firms (known as “original equipment manufacturers” or OEMs[2]) that regularly develop sophisticated products and systems and the entrepreneurial firms that engage in prototyping or pilot production to advance new products.

This chapter focuses on opportunities for building innovation capacity within the Massachusetts manufacturing ecosystem and, in particular, on how the state can best support SMEs in their efforts to be globally competitive. Manufacturing capabilities are grounded in particular regions, where, historically, they have grown around key industries—examples include the automotive industry in the Midwest or turbine engines and firearms in Massachusetts. Thus, manufacturing lends itself to regional approaches for increasing innovation capacity and upgrading firms’ capabilities. Strengthening the regional innovation ecosystem as a whole will improve the “industrial commons” [34] and leverage greater results by helping all manufacturers in the state, not just a select few.

This is particularly important for SMEs. Recent research by MIT’s Production in the Innovation Economy (PIE) project [4] concluded that SMEs often find themselves “home alone” when it comes to competing globally and driving innovation in their companies. The large, vertically-integrated corporations of the 1980s have tended to become less vertically integrated over time as they sought to focus on their core competencies, outsourced much of their production and increasingly relied on smaller suppliers to drive innovation. This process has left “holes” in the industrial ecosystem, such as diminished innovation and R&D support, cutting off many of the important investments and spillovers that used to flow from large corporations to smaller firms (e.g., in training, technology adoption, and R&D investments). As a result many SMEs have been left largely on their own to figure out how to find and train workers, adopt new technologies, and develop and scale new products and services, while shouldering the burden of funding this at the same time.

This chapter focuses on how to fill these holes as they relate to innovation. For this, a systems approach is used that considers how knowledge and sources of innovation flow between key participants within the manufacturing innovation ecosystem. Strengthening these links and expanding the flow of knowledge between key actors will upgrade the system as a whole and enhance the region’s competitiveness. As other regions and countries around the world increase

[1]Small and medium-sized enterprises (SMEs) refer to firms with fewer than 500 employees. Interestingly, the U.S., unlike Europe, does not use revenue to define SMEs [39].

[2]Original equipment manufacturers (OEMs) are “firms that […] manufacture […] based on ‘original’ designs” [37]. OEMs either make products directly or act as a system integrator before selling directly to the customer. Throughout this chapter, the term OEMs typically refers to large enterprises, with over 500 employees.

investment in manufacturing and incentives for manufacturing firms, it is increasingly important for Massachusetts to leverage and invest in its own innovation assets.

2 Definition of Key Terms

Discussing Cyber-physical Production Systems key terms concerning this concept need to be defined. Figure 1 gives an overview of the relation of those key terms in a nutshell.

The Internet of Things (IoT) is a global infrastructure enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies [17]. A special manifestation of this is the Industrial internet that is an internet of things, machines, computers, and people, enabling intelligent industrial operations using advanced data analytics for transformational business outcomes [16].

In addition to that, the term Cyber Physical Systems (CPS) is frequently used in this realm. CPS are embedded systems that are connected to each other using digital networks and have several multi-modal human-machine interfaces and gather and process physical data through sensors and interact with physical operations through actors [1]. The core of such CPS are embedded systems which are microprocessor-based systems that are built to control a (range of) function and are not designed to be programmed by the end user [14].

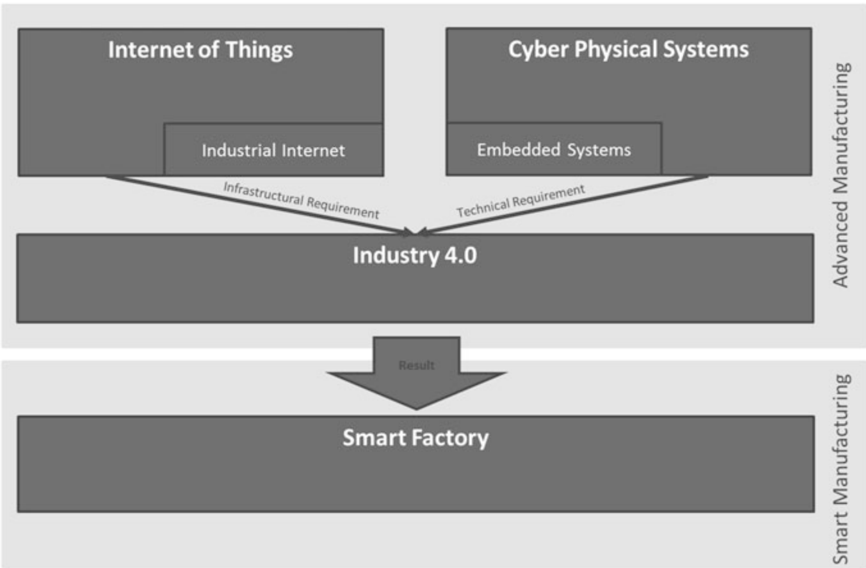


Fig. 1 Key terms in advanced manufacturing

Both IoT and CPS are the infrastructural and technical requirements respectively for the proclaimed fourth industrial revolution or Industry 4.0 [10].

The term Advanced Manufacturing comprises all those aspects although in a very broad sense. It refers to the use of next-generation technologies in manufacturing processes. Specifically, advanced manufacturing "makes extensive use of computer, high precision, and information technologies integrated with a high performance workforce in a production system capable of furnishing a heterogeneous mix of products in small or large volumes with both the efficiency of mass production and the flexibility for custom manufacturing in order to respond rapidly to customer demands" [18]. More precisely, advanced manufacturing encompasses "a family of activities that depend on the use and coordination of information, automation, computation, software, sensing, and networking, and/or make use of cutting-edge materials and emerging capabilities enabled by the physical and biological sciences. It involves both new ways to manufacture existing products and the manufacture of new products emerging from new advanced technologies" [32].

The result of advanced manufacturing or Industry 4.0 is Smart Factories or Smart Manufacturing. The former refers to factories with decentralized manufacturing logic through intelligent products in horizontally and vertically integrated production systems for a consistent engineering along the value chain [10]. The latter means the same but without explicitly mentioning the decentralized control and can be seen as a sophisticated practice of orchestrating the use of data-driven manufacturing intelligence using multiple real-time SM Systems pervasively deployed throughout all operating layers across the entire factory and networks [9].

It is worth beginning by defining a term that is frequently used and often poorly specified: innovation. According to one source, innovation refers to "the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organizational method in business practices, workplace organization or external relations" [4]. Innovation differs from invention. Invention is the creation of something new and novel while innovation is the process of adding value to an invention such that it becomes useful in the marketplace [5].

There are four different dimensions to innovation (Fig. 2). Product or service innovation is the first-time commercial utilization of a product or service that is new to the market, whereas process innovation is the implementation of methods that are new to the company, but not necessarily new in the market, and that change the way a company manufactures a product. Process improvement measures, like lean manufacturing, Six Sigma, etc., are often included in this category of innovation, though they may be less about true innovation and more about continuous improvement. Furthermore, organizational innovation is the implementation of new organizational methods within a firm that change the firm's business practices, communication, and/or workplace organization. The latter two innovation dimensions have a clear company perspective. In the following, the primary focus will be on product and process innovation.

Innovations are only realizable if embedded in a fruitful ecosystem. "Innovation ecosystem" is a term that has gained popularity in recent years. The "ecosystem" metaphor draws from our understanding of natural ecosystems and of their ability to

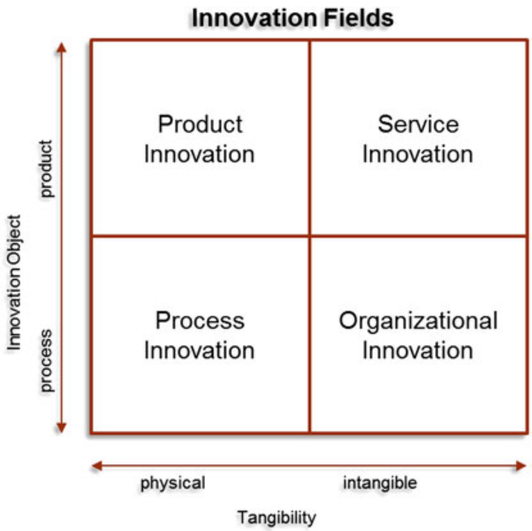


Fig. 2 Dimensions of innovation [21]

sustain a population when all members of the community are contributing. The idea of an "innovation ecosystem" is rooted in part in the literature on "national innovation systems" [24]. A national innovation system (NIS) is defined most succinctly as "the set of institutions whose interactions determine the innovative performance of national firms." The term "ecosystem" adds a more dynamic element to the system concept [29].

3 Trends in Advanced Manufacturing

The marriage of hardware and software, and the use of new information technologies combined with advanced machinery to increase automation, intelligence, efficiency, and sustainability in manufacturing processes is at the heart of recent developments in advanced manufacturing. In 2013, Germany launched its "Industry 4.0" initiative with a primary focus on the systematic interconnection of existing manufacturing systems in the new "facility of the future" (machinery, information systems, employees, regulation, standardization) [10] in order to develop self-organizing autonomous manufacturing systems. In the U.S., the Smart Manufacturing Leadership Council has adopted a slightly different emphasis, albeit with similar goals. The Council is focusing on the need to develop new standards and platforms for a common information technology infrastructure that would include, for example, data collection systems and community simulation platforms [36] for new advanced manufacturing technologies.

More broadly, in two reports to the President's Council of Advisors on Science and Technology (PCAST) in 2012 and 2014, the Advanced Manufacturing Partnership (AMP), an industry-academia-government partnership, put forward several recommendations for boosting innovation in advanced manufacturing in the U.S.

through the creation of new R&D infrastructure and technology road maps. The National Network for Manufacturing Innovation (NNMI), which was launched in 2012, represents the country's most significant investment in advanced manufacturing in recent history. It includes several centers that are supported and led by public-private consortia and that focus on the development of pre-competitive technologies while also building regional capabilities in their focus areas [3]:

- Institute for Advanced Composites Manufacturing Innovation (Knoxville, TN)
- Digital Manufacturing and Design Innovation Institute (Chicago, IL)
- Lightweight Innovations for Tomorrow (Detroit, MI)
- PowerAmerica—Wide Bandgap Semiconductors (Raleigh, NC)
- America Makes—Additive Manufacturing (Youngstown, OH)
- Flexible Hybrid Electronics (San Jose, CA)
- Integrated Photonics Institute (Albany, NY)
- Revolutionary Fibers and Textiles (in progress)
- Smart Manufacturing (in progress)

The recent AMP 2.0 report in 2014 highlighted three additional focus areas for future national efforts in manufacturing innovation:

1. advanced sensing, control and platforms in manufacturing;
2. visualization, informatics and digital manufacturing; and
3. advanced materials manufacturing.

Several national studies are relevant to this report, including the aforementioned AMP reports to PCAST in 2012 and 2014 [33], MIT's 2013 Production in the Innovation Economy study [4], and a recent (2015) report on supply chains by the U.S. Department of Commerce [12], among others. These national reports address a range of important issues such as enabling innovation, improving training and the talent pipeline, strengthening supply chains, and generally rebuilding the industrial ecosystem while improving the overall business environment.

At the local and regional level, recent reports specific to Massachusetts focus primarily on the needs of SME manufacturers and on the state's business environment. They highlight the high need for more skilled workers, the cost of doing business, the need for technical assistance and innovation support, the importance of access to capital, and the value of a better image for manufacturing. Although these regional studies provide detailed information about the manufacturing base in Massachusetts, a comprehensive analysis of the manufacturing innovation ecosystem has not been the focus of regional work to date.

4 The Competitive Position of Manufacturing in Massachusetts

Massachusetts offers an important case study of how small U.S. manufacturers compete in today's global economy and complex supply chains. The Commonwealth has a diverse and sophisticated manufacturing base that includes about 7,000 firms in a wide range of industries, including aerospace/defense, semiconductors/electronics, medical devices, and biopharmaceuticals [7]. SMEs with fewer than 100 employees account for about 92 % of the manufacturers in the state [38]. The vast majority of these firms participate in global supply chains. However, SMEs account for only approximately 30 % of the state's manufacturing employment [31].

Massachusetts has a long and illustrious history in manufacturing and in product and process innovation [15], and the advanced manufacturing capabilities it built over the past 150 years have allowed companies and workers to transition into new or emerging industries as market conditions change. In fact, one of the region's strengths is a diverse manufacturing base that supports cross-fertilization among its key clusters.

Manufacturing employment has steadily declined over the past several decades (see e.g., [7]). Since the 1990s manufacturing jobs have declined as a share of the state's overall employment from approximately 19 % to about 9 % today (this compares with a national-level figure of about 11 % in 2013, down from 20 % in 1990 [5]), where the current data reflect some recovery from the depths of the Great Recession in 2008. The decline in the share of manufacturing jobs at the state level mirrors national trends for the U.S. as a whole, and global trends for other industrialized countries as productivity rates have increased, production has become more fragmented, and global competition has intensified.

While Massachusetts manufacturers are undoubtedly operating in an increasingly complex environment, this new environment also offers opportunities for those SMEs who can compete on a "world-class" basis. More intense global competition, the development of new generations of advanced manufacturing technologies, and novel ideas about how to organize manufacturing firms and facilities and better deploy workers are creating challenges and new possibilities for advanced manufacturing SMEs.

Despite the fact that a significant number of Massachusetts SMEs are engaged in contract manufacturing of what are often termed (misleadingly) "commodity products," OEMs consistently referenced the following attributes as key characteristics of the state's manufacturing production system:

- Small-batch niche production, rather than large-volume mass production;
- Extremely high quality and performance requirements (0 % failure);
- High knowledge and innovation content;
- New or early-stage products and prototyping;
- Products with high proprietary content;
- Products where proximity to market is desirable;
- Products where regulatory factors encourage siting in the U.S.; and
- Customized products with quick turnaround time if needed.

To sustain these characteristics, empirical research showed that OEMs can draw on four primary assets:

- A well-educated and highly skilled labor force, particularly in engineering;
- Suppliers that are able to quickly provide difficult-to-manufacture parts of very high quality and reliability;

- World-class universities; and
- Innovative startups and a dynamic entrepreneurial ecosystem.

For all these reasons, Massachusetts manufacturing base has stabilized since the 2008 crisis and remains strong today. Indeed, the state's manufacturers are well positioned to take advantage of some of the national and global trends that suggest the U.S. may be more globally competitive in manufacturing in the future. In particular, declining energy costs, rising labor costs in traditionally low-wage countries, and concerns about the protection of intellectual property are making the U.S. environment more competitive for certain types of manufacturing, including those in which Massachusetts excels. In addition, the development of new "game-changing" advanced manufacturing technologies such as additive manufacturing, cyber-physical systems, and integrated circuit photonics, is providing additional opportunities for U.S. firms to innovate and increase efficiency.

Manufacturing in Massachusetts was adversely affected by the recessions of 2000 and 2008, which caused the state to lose 40 % of its manufacturing employment base and 30 % of manufacturing establishments (Figs. 3 and 4). In terms of employment, Massachusetts followed the national trend with a sharp decline in manufacturing jobs in 2008 and 2009, followed by a stabilizing of the employment picture in 2010 at approximately 250,000 workers and 7,000 establishments.

In terms of manufacturing establishments, Massachusetts experienced a steady decline from about 10,000 establishments in 2001 to about 7,000 in 2013, for a total contraction of about 30 %—twice the national rate. The total number of U.S. manufacturing establishments fell from about 400,000 in 2001 to about 335,000 in 2013 (Fig. 4).

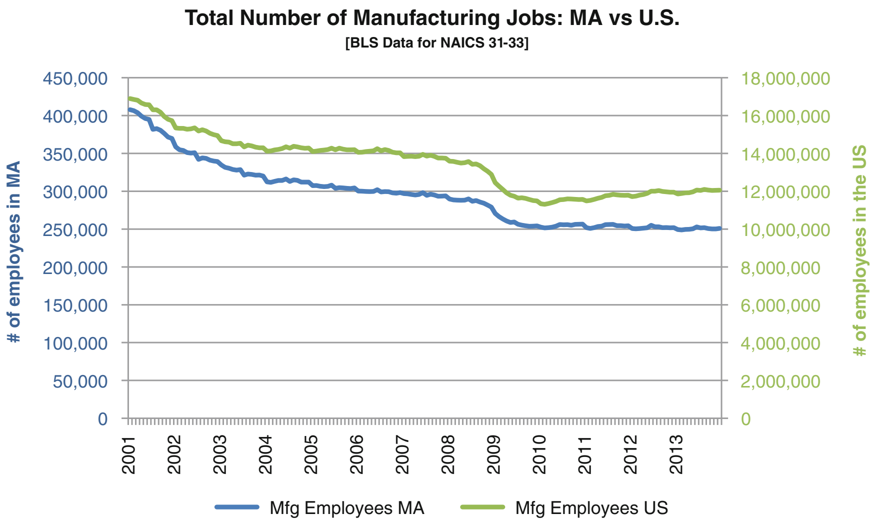


Fig. 3 Total number of jobs in the manufacturing industry in Massachusetts and in the United States between 2001 and 2013 [6]

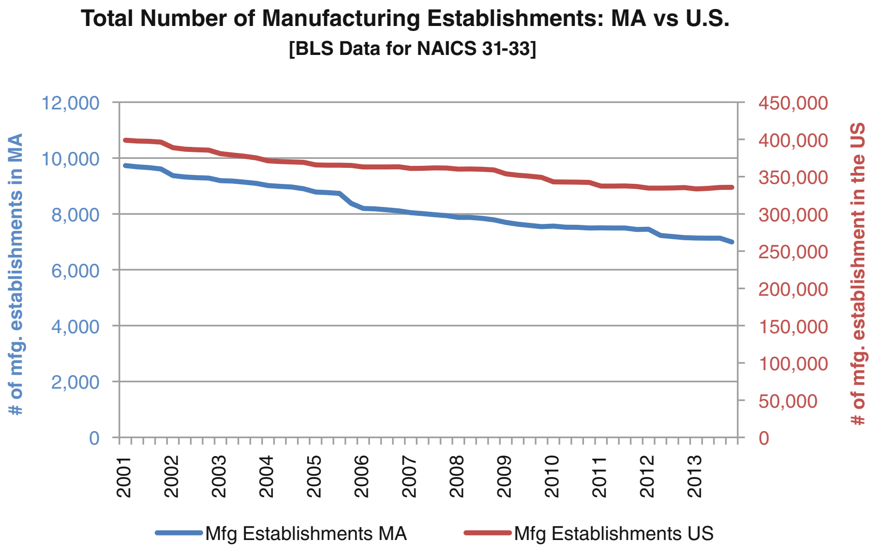


Fig. 4 Total number of establishments in the manufacturing industry in Massachusetts and in the United States between 2001 and 2013 [6]

Approximately 97 % of all manufacturing establishments in Massachusetts can be considered SMEs (i.e., firms with fewer than 500 employees) [31] and about 92 % have even fewer than 100 workers [38]. Although SMEs vastly outnumber large firms, they account for only 30 % of all manufacturing jobs. Large firms, though they represent only approximately 3 % of all manufacturing establishments, account for 70 % of manufacturing employment in the state [31].

5 The Massachusetts Manufacturing Innovation Ecosystem

The innovation process is often characterized as non-linear and dynamic, involving different actors with highly interactive relationships [20, 41]. While firm innovation might have occurred in isolation in the past, particularly when many firms were vertically integrated, today's firms have high degrees of interaction with a range of other companies and organizations, such as universities, suppliers, customers, and even competitors—all of which may play a part in building a firm's innovation capacity. External factors such as laws, regulations, culture, and technical standards also play an important role in setting the stage for innovative activities [11]. For these reasons, the process of innovation cannot be viewed through one single lens (within a single company or institution) but needs to be understood as part of a larger system [26, 40]. This is the approach taken in the following.

Based on our research, four key nodes and associated institutions and actors play a major role in the state's advanced manufacturing innovation ecosystem:

- Large OEMs,
- Supplier SMEs,
- Startups, and
- Universities and research institutions.

Figure 5 presents a stylized representation of these key drivers and actors. Obviously, the innovation system relies not only on flows between the four nodes depicted in the figure but also on knowledge that comes into the region from outside sources such as R&D networks, trade associations, and global partnerships/networks.

The lines connecting each of the four nodes represent the general strength and direction of the knowledge flows between them.

In general, OEMs have the strongest links within the innovation ecosystem because they are largely driving innovation activities within it. Knowledge flows between OEMs and research universities are strong in both directions, while knowledge flows with SMEs are relatively unidirectional flowing from OEM to SME. With respect to innovation, startups typically bring new ideas to the OEMs.

In contrast to OEMs, SMEs generally have the weakest links within the ecosystem. Historically, they have most often been on the receiving end of knowledge flows from their large customers. Their ability to drive knowledge and ideas in the other direction, toward the OEMs, has been limited, though this is highly dependent on the OEM. SMEs also generally have weak links to universities and to the startup community.

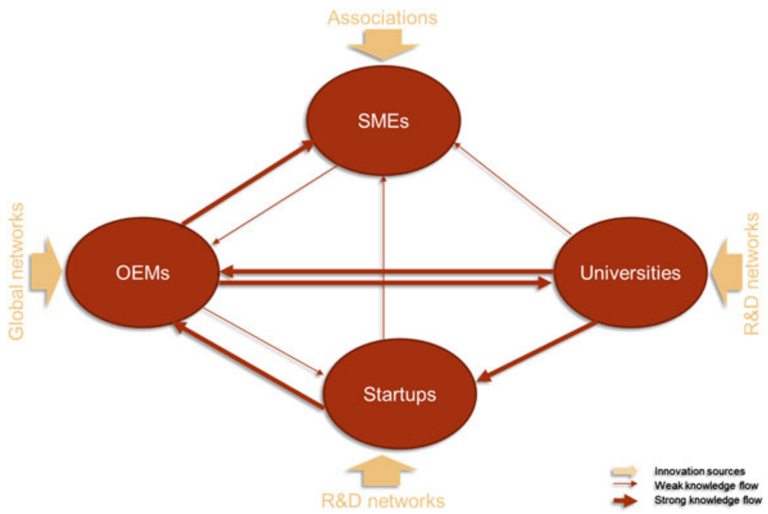


Fig. 5 A schematic of the manufacturing innovation ecosystem in Massachusetts

Universities have relatively strong links with large OEMs and with the startup community, but limited engagement with SMEs. University research primarily drives "disruptive" innovation and is often focused 10–15 years out in terms of new technological developments.

Finally, the vibrant startup community is an important source of innovation for OEMs. The strength of the link between startups and OEMs depends in part on the industry and on the extent to which OEMs are receptive to, and actively engaged with, the startup community.

5.1 OEMs Within the Manufacturing Innovation Ecosystem

OEMs are the most important drivers of innovation in Massachusetts with connections to all other actors in the innovation ecosystem. Interviews with OEMs in the top 10 manufacturing sub-industries[3] suggested that OEMs draw on the region's capabilities in different ways depending on their industry structure, their development time horizons, and their regulatory environment. In all cases, OEMs consider the region a place for new product development and new product introduction as evidenced by the number of OEM advanced manufacturing R&D facilities located in the state (some company examples include Gillette, Medtronic, Thermo-Fisher, Raytheon, and more recently Nihon Kohden and Phillips Healthcare). For example:

- Semiconductors and electronics are largely manufactured in Asia and Mexico and then integrated into other products in the U.S.; there is some specialized production in the U.S. as well.
- The aerospace and defense industries require largely domestic production, but there is increasing pressure on OEMs to manufacture in the countries of their foreign customers.
- Manufacturers of measuring devices and medical devices are more likely to keep high-end production in the U.S.; they benefit from proximity to suppliers for rapid response and small-batch production.

As described earlier, OEMs manufacture in Massachusetts for reasons largely linked to innovation and talent. Access to innovation and talent helps the OEMs respond to increasing pressure to cut lead times and meet high quality standards.

[3]We used three filters to help determine which manufacturing sub-industries could be considered especially advanced or innovative, starting with NAICS (North American Industry Classification Systems) codes at the four-digit level, we considered (1) patent data as a proxy for innovation, albeit one that is not particularly well suited for manufacturing; (2) R&D spending per worker and share of STEM (science, technology, engineering, and math) occupations; and (3) employment data. The identified sub-industries this way are Analytical Laboratory Instruments, Detection and Navigation Instruments, Process Variable Measuring Instruments, Semiconductors Machinery, Semiconductors and Related Devices, Electronic Computers, Aircraft Engines, Medical Instruments, Pharmaceuticals, and Machine Shops.

Interviews highlighted the following attributes of the Massachusetts innovation ecosystem:

1. The presence of world-class research universities with high-impact research groups gives OEMs the opportunity to support unique, business-related, cutting-edge research that can be integrated or translated into competitive products to gain market share.
2. Graduates from the state's research universities constitute an important talent pool for large OEMs as they seek to develop new or improve existing products, services, processes, or organizational structures.
3. Besides universities, the state's vibrant startup community is a source of new ideas for products and services; in addition, collaboration with or acquisition of business-related startups can open new market opportunities.
4. To rapidly introduce new products, OEMs in Massachusetts can rely on flexible, quick, and reliable suppliers, especially machine shops that can manufacture special parts and components on a small scale.

At the same time, OEMs in Massachusetts face several key innovation challenges:

1. While a well-educated, highly skilled labor force is one of the Commonwealth's major strengths, OEMs are emphatic that access to labor remains a serious problem. This is an area where Massachusetts is under strong pressure from other regions. Several OEMs expressed the view that the supply of labor—including skilled labor—was better in the South, and in some cases better abroad (especially in Mexico).
2. The younger generation's perception of manufacturing jobs is out of date and needs to be updated to reflect the clean, technologically advanced nature of the industry. The Massachusetts manufacturing community is acutely aware of the problem of skilled labor shortages and has taken a number of actions in response, including strengthening its outreach to community colleges and local organizations to promote manufacturing as a viable career, and revising and standardizing training programs to facilitate skills acquisition.
3. The importance of government's role in attracting or retaining manufacturing investments cannot be ignored. Some U.S. states have taken a very aggressive approach in trying to attract manufacturing jobs, actively recruiting manufacturing firms and offering significant incentives to locate manufacturing facilities in their state. Further, governments of many developing or emerging economies (e.g. South Korea, Turkey, Brazil, the Middle East) require suppliers to set up operations in the country if they would like to do business there. U.S.-based OEMs have often responded to such requests without moving essential manufacturing but these kinds of quid pro quo or offset pressures are increasing. Finally, several OEMs perceive that China is progressively losing its attractiveness as a low-cost manufacturing location because of rapid wage escalation, poor workforce stability, and the total costs of addressing intellectual property protection.

Several important efforts are already underway in Massachusetts to address the issue of labor supply and training and to begin changing perceptions about the nature of manufacturing jobs. The Manufacturing Advancement Center Workforce Innovation Collaborative (MACWIC), for example, is collaboratively tackling urgent issues, like workforce training. The MACWIC program is employer-led; it comprises not only companies, both small and large, but also education/technical training providers as well as the Massachusetts Manufacturing Extension Partnership (MassMEP) and it aims to identify and find solutions to workforce-related needs. The Collaborative's most important output to date is a five-tiered training pyramid consisting of stackable consecutive training modules that can be offered jointly by vocational and technical high schools and community colleges. Students can take these modules to earn an Associate Degree in Manufacturing Technology. The industry-driven and modular nature of the program enables employers to better evaluate the level of graduates' skills [25].

5.2 SMEs Within the Manufacturing Innovation Ecosystem

As noted earlier, 97 % of all manufacturing establishments in Massachusetts are small or medium sized. Machine shops account for a significant number of these manufacturing SMEs; typically, they perform contract manufacturing and work within regional, national or international supply chains.

The SME landscape in Massachusetts includes four different types of businesses. These are classified according to company life-cycle and type of product architecture, as depicted in Fig. 6. Along the horizontal axis, the figure distinguishes between newly founded and incumbent SMEs; along the vertical axis, the distinction is between SMEs that produce parts and components and SMEs that make end products.

Startup or spin-off suppliers produce less complex parts and components and seek to engage with large OEMs to sell their products. In terms of life cycle, high-performing startup or spin-off suppliers are on a path to grow to mature small suppliers (the fourth quadrant in the figure) and ultimately to become strategic suppliers.

Small suppliers normally start as emerging startup or spin-off suppliers and grow to become part of OEM supply chains for precision parts. The large number of machine shops in Massachusetts fit into this category. Machine shops are not positioned to become strategic partners because they are engaged in made-to-order manufacturing of less complex parts and do not have proprietary products. For these suppliers, support to improve process efficiency with initiatives like lean manufacturing or Six Sigma is essential.

Startup or spin-off OEMs produce more complex, proprietary products that can be marketed by the OEM or as part of a larger system. The pathway for these kinds of SMEs is to grow through new customers and markets into a mature small OEM and ultimately to become a large OEM.

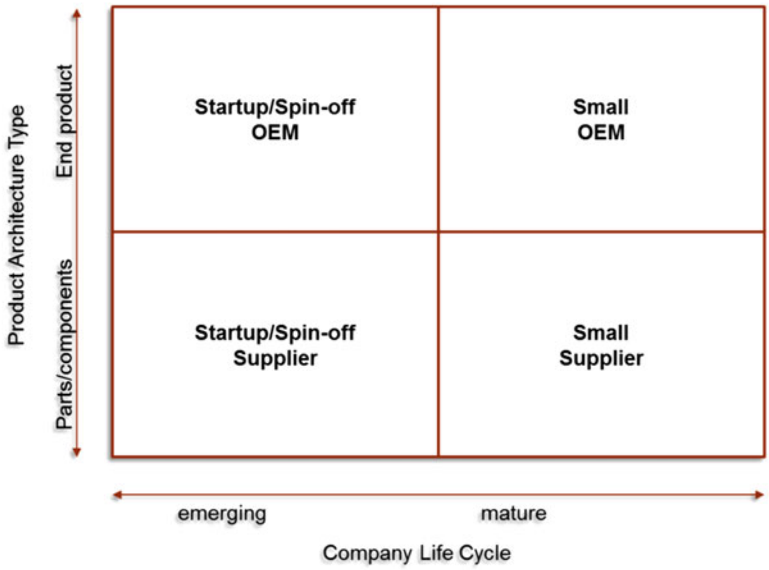


Fig. 6 Classification of SMEs

Small OEMs, which have their own product portfolio, seek to enter new markets and connect with other OEMs. These companies are also often well positioned to partner with universities.

We interviewed several small suppliers and small OEMs that could be considered high performing, or on the way to becoming high performing, for this research. A number of precision engineering firms (referred to more generally as machine shops) were included in this group. These SMEs often enable new product introductions and undertake prototyping activities for OEMs based in the region. They also play important roles as providers of key equipment and strategic parts for OEMs' physical production systems within the state.

Several existing and potential pathways are available to increase the innovation capacity of SMEs in the state.

Despite growing cost pressures and increasing consolidation within OEM supply chains, Massachusetts SMEs are not only in a good position to take advantage of heightened interest in innovation and shorter lead times, they may also be buoyed by trends that are making manufacturing in the U.S. more attractive generally. OEM interest in greater collaboration also creates new opportunities to build long-term relationships.

In addition, the relatively diverse manufacturing-related key sectors of the Massachusetts economy that rely on the state's "manufacturing backbone" provide a diverse customer base for SMEs. The ability to supply across sectors helps SMEs in terms of business cycles, cross-selling, and also cross-fertilization with respect to learning and best practices.

Strong institutional support is available in Massachusetts for process improvements and workforce training to help SMEs produce more efficiently.

In terms of challenges within the innovation ecosystem, a primary challenge for SMEs is that they are not easily integrated into structures for learning about and participating in the development of new products and processes, including frontier technologies. Access to this knowledge, whether from OEMs or universities or other third parties, is limited. In particular, despite some pilot efforts within the state, SME relationships with universities are weak. In interviews, SME managers repeatedly stated that many universities are not "user-friendly" places—that is, they are frequently hard to navigate.

Weak linkages with startups are a further challenge for SMEs. Improving these linkages could open new market opportunities, especially since the vibrant startup community in the greater Boston area needs manufacturing services that could be delivered by small suppliers such as machine shops.

5.3 Universities in the Manufacturing Innovation Ecosystem

Much has been written about the important role universities play in fostering innovation and generating economic development benefits for the regional economies in which they operate. One of the obvious advantages of a university to the ecosystem is that, "unlike so many participants in the local economy, they are immobile" [23].

Massachusetts universities in particular have an enormous impact on the region's economy: throughout the state, about 500,000 students are enrolled in more than 100 institutions of higher education and billions of dollars go to support world-class basic and applied research at these institutions. Entrepreneurial activities on university and college campuses have led to the founding of many innovative startup firms.

In addition to all the tangible outcomes they generate, universities also create many positive externalities for surrounding communities [2] and play at least two important roles that can help foster regional economic development. First, universities create a "space for open-ended conversations about industry development pathways and new technological and market opportunities." They also "increase the local capacity for scientific and technological problem-solving" through the flow of ideas from startups, joint research with companies, consulting, and the hiring of students [23].

Advanced manufacturing in Massachusetts has benefited from all of these innovation externalities associated with local universities and colleges. In particular, the state's universities boast top research labs and centers (often supported in part by state and federal funding) that are developing the next generation of advanced manufacturing technologies. Examples include the recently launched Raytheon-UMass

Lowell Research Institute, which is focused on flexible and printed electronics and the Novartis-MIT Center for Continuous Manufacturing, which focuses on biomanufacturing. Both of these centers are sponsored by large OEMs and support basic and applied R&D. Other centers build on regional strengths in areas such as robotics (e.g., the Wood Hole Oceanographic Institution Center for Marine Robotics and the UMass Lowell New England Robotics Validation and Experimentation Center or NERVE), advanced materials (e.g., the MassNanoTech Institute at UMass Amherst, the Northeastern Nanoscale Technology and Manufacturing Research Center, and MIT. Nano), life sciences (e.g., the MIT Medical Electronic Device Realization Center or MEDRC and the UMass Lowell Biomanufacturing Center), defense-related research (e.g., Draper Labs and the U.S. Army Soldier Research, Development and Engineering Center), and advanced manufacturing technologies more generally (e.g., the Advanced Technology and Manufacturing Center at UMass Dartmouth, the Lab for Manufacturing and Productivity at MIT, and the Fraunhofer Center for Manufacturing Innovation at Boston University). Some centers, like the UMass Dartmouth Massachusetts Accelerator for Biomanufacturing, are designed specifically to work with startups that can benefit from the use of shared facilities.

Clearly, universities are already a critical part of the state's manufacturing innovation ecosystem. Moreover, they are positioned to play an even greater role going forward given the current focus on emerging technologies and industries that are important to the Massachusetts economy.

We identified at least two areas of opportunity for universities in the state's advanced manufacturing innovation ecosystem.

Despite flat or declining public funding for basic research in recent years [30], advanced manufacturing has attracted significant national attention and investment. The creation of a National Network for Manufacturing Innovation (NNMI) [3], which proposes to create at least 15 Institutes for Manufacturing Innovation (IMI) around the country (see above), is arguably one of the most important science and technology initiatives put forth by the federal government in recent years. This effort recognizes the importance of manufacturing to the country's innovation capacity and is based in part on the German Fraunhofer Institute model and the applied research model of public/private, university and large/small company collaborations.

Massachusetts universities have submitted bids in response to NNMI last calls for proposals and will no doubt be included in future bids, given the range of expertise that exists in the state. While the hope is that Massachusetts will host at least one IMI,[4] the extensive work and initial collaborations that have been prompted by the NNMI process will yield benefits even if no Massachusetts institution is successful in the national competition. Those involved with this process should convene to discuss what aspects of individual bids could be implemented at the state level, potentially building synergies across bids.

[4]Thus far, Massachusetts' universities, esp. MIT, are participating in three IMIs.

The NNMI process could also be helpful in terms of developing advanced manufacturing technology road maps for the region. Such road maps would identify advanced manufacturing technologies of particular importance to the state's leading industry clusters and develop ideas for how best to support and advance research in these areas.

Another area of opportunity for strengthening the manufacturing innovation ecosystem involves increasing the engagement between universities and SMEs. While there have been some successful examples and pilots, some fundamental obstacles exist that make such collaborations challenging.

First, SMEs face organizational challenges when working with universities. As already noted, SMEs report that they find universities hard to navigate and not user-friendly. Second, universities and SMEs have different objectives and agendas. Academics see innovation as "something that is radically new deriving from newly created knowledge" while SMEs see innovation as creating a product or process that will increase the firm's profits [27]. Third, SMEs are usually working under short- or medium-term time constraints. Universities work with longer timeframes. Finally and perhaps most importantly, the costs of collaboration can be prohibitive unless funding is provided by the SME or a third party.

Finding ways to engage SMEs in research and discussions about new technologies is crucial to increasing their innovation capacity. One way to engage SMEs in university collaborations is through competitive grants, like those offered by the Small Business Technology Transfer (STTR) program. Facilitating and broadening SME-centered industry-university collaborations offers another promising path for increasing innovation capacity among SMEs [19].

5.4 Startups in the Manufacturing Innovation Ecosystem

Massachusetts is widely regarded as one of the most innovative and entrepreneurial states in the country.[5] Innovative startups, which may grow out of universities or out of larger established firms, are at the heart of the state's innovation ecosystem.

What is less well known is the extent to which these startups are engaged in advanced manufacturing processes. Research on startups based on technology developed at MIT and licensed through the MIT Technology Licensing Office (TLO) found that approximately 80 % of all TLO startups founded between 1997

[5]The Milken Institute's State Technology and Science Index 2014 as well as the ITIF's 2014 State New Economy Index rank Massachusetts as number one. The former analyzes technology and science capabilities of each U.S. state alongside their success at transforming those capabilities into companies [8]. The latter evaluates states' fundamental capacities in the "new economy" "in terms of knowledge jobs, globalization, economic dynamism, digital economy, and innovation capacity" [13].

and 2008 required some kind of production-related capabilities [35].[6] In addition, a study of Massachusetts firms that are receiving federal Small Business Innovation Research (SBIR) grants found that at least 15 % (or 500 firms) that received grants between 2009 and 2013 were engaged in advanced manufacturing processes. These grants accounted for approximately $200 million of the $1.2 billion total that Massachusetts firms received in SBIR grants over this time period.[7]

Given the region's strong and growing engineering capabilities and the trend toward combining hardware and software to form "hybrid" technologies (in consumer and medical devices, for example), startups have become an increasingly important source of manufacturing innovation. The emergence of startup incubators/seed funds such as Bolt that focus on hardware companies reinforces the support system for such startups.

But startups also face challenges in the scale-up phase. Growing innovative companies is a subject that is increasingly drawing attention, both in the United States and globally, as regions and countries focus on reaping some of the downstream benefits of their startup ecosystems. The scale-up process is particularly challenging for startups engaged in the production of complex production-oriented technologies (as opposed to software). Such technologies often require larger amounts of capital and longer time horizons (often over 10 years) to demonstrate their viability at commercial scale [35].

There are several points in the early stages of this process where actively engaging with the manufacturing innovation ecosystem could help startups achieve scale and, importantly, facilitate scale-up in the Commonwealth.

First, startup technology companies often have a promising idea for a new product but lack the skills to manufacture it. Early-stage prototyping, which requires multiple iterations that can take several months or several years, often requires close proximity between the startup and its suppliers so that the latter can respond to changes quickly while still providing high quality. Massachusetts, with its extensive network of high-precision machine shops and experience in new product introductions, provides competitive advantages to startups at this stage of development. However, connections between the innovative startup community and the state's high-precision machine shops are weak, with few formal or systematic forms of interaction. One manager of a startup suggested that companies in the greater Boston area are potentially as likely to connect with suppliers in California or China as they are to connect with suppliers in Massachusetts. Thus, it will be important to underscore the region's capabilities in prototyping and early-stage piloting and open better channels of communication between these communities. A recent pilot with Greentown Labs, an incubator for clean energy companies, exemplifies a first step in this process. Whether the state can also position itself to

[6]Generally speaking, firms that license technology through the TLO are less likely to be software-related.

[7]In terms of total SBIR and STTR grants, Massachusetts is the second most successful state in the country behind California and is the leading state in the country in terms of SBIR/STTR grants per capita.

support scale-up beyond pilots remains to be seen. Recently, companies have been more likely to go abroad to lower-cost locations for commercial scale-up.

A second area of opportunity for supporting the scale-up process in the region is with potential customers. Early adopters are among the most important factors that can help a startup "cross the chasm" in the early stages of scale-up [28]. Customers or potential customers who are willing to partner during beta testing of a new product are critical. Increasingly, strategic partners have been playing this role in the United States. Such partners, which are usually large companies (including OEMs), are becoming more engaged in startups through equity investments and other arrangements [22] in which they provide not only capital but capabilities and know-how in exchange for the exposure and experience they gain from the startup. This is particularly important for startups that, because of their longer development horizons and higher capital needs, do not necessarily fit well with a venture capital funding model. Given the diversity and sophistication of OEMs in Massachusetts, a more systematic effort could be made to connect startups and OEMs through e.g. intermediary-led forums, match-makings, etc. This would benefit both parties as well as the regional economy. Introducing large potential customers to startups is the goal of several initiatives that are already in place (e.g., the NECEC Strategic Partners program and Fintech Sandbox's efforts to provide scrubbed financial data from large financial services firms to financial services startups for beta testing). More could be done in this area, particularly with respect to advanced manufacturing companies, where the scale-up process can be more challenging due to capital requirements and longer time horizons.

This review of the four key nodes in the manufacturing innovation ecosystem—OEMs, SMEs, universities, and startups—highlights the multiple ways these actors coexist within the same regional innovation ecosystem, often working closely together, but in some cases missing opportunities for greater collaboration and greater overall enhancement of the region's innovation capacity.

6 Manufacturing Intermediaries

Massachusetts is rich in intermediaries that provide, among other things, services and advice to SME manufacturers throughout the state. This assistance takes six primary forms:

1. process improvements,
2. workforce training,
3. strategic technology and cluster development,
4. technical and engineering process support,
5. managerial and professional education, and
6. marketing.

However, the current system tends to focus on “point solutions”—such as supporting SMEs on a one-on-one basis primarily in workforce training, lean practices, and certification. This is necessary but not sufficient in terms of building innovation capacity. State efforts to support SMEs also focus primarily on the supply side—i.e., on workers and suppliers—often without enough input from the OEMs that drive the demand side. In addition, despite investments in some emerging technologies, Massachusetts lacks an overall strategic vision for advanced manufacturing that looks out 5–10 years in terms of supply chain developments, technology road maps, and talent and training needs.

7 Recommendations to Improve the Innovation Ecosystem

Based on these findings, we identify four distinct areas of opportunity for improving the Massachusetts manufacturing innovation ecosystem, particularly for SMEs. They involve a statewide manufacturing strategy and agenda, OEM collaboration, technological and managerial support, and connections with startups. Our recommendations in each of these four areas are summarized below.

7.1 Advanced Manufacturing Strategy and Agenda

- **Develop an Advanced Manufacturing Strategy for the State**

In contrast to the state’s other cluster-focused strategies (e.g., for the biotech industry), advanced manufacturing requires the development of cross-cutting capabilities that work across industries. This makes it more challenging to develop strategies around particular capabilities. A deep understanding of advanced manufacturing capabilities, their importance within key clusters, and trends in technology as well as in the global manufacturing marketplace is required.

A robust analysis of the state’s advanced manufacturing capabilities combined with engaging key manufacturing leaders in the state is necessary to develop an advanced manufacturing strategy and agenda for the next 5–10 years. This includes involving relevant stakeholders and establishing appropriate governance structures to oversee such an effort.

- **Introduce Consortium-Based Applied Research Projects**

Grant funds should be used to encourage regional consortium-based projects including Universities, OEMs, and SMEs that focus on pre-competitive product and process innovations, similar to the German model. Experience in consortium-building in the process of applying for the federal Institutes for Manufacturing Innovation (IMIs) could be instructive in developing regional, project-based consortia.

7.2 Collaboration with OEMs to Drive Innovation and Upgrade SME Capabilities

- **Inaugurate a Commonwealth Manufacturing Innovation Advisory Group**

OEMs are a driving force for innovation in Massachusetts, yet their collective voice on the subject is not being heard. With a window into global trends, R&D opportunities, supply chain demands, and training needs 5–10 years out, OEMs need to be engaged in helping set the state's manufacturing innovation strategy going forward. Their participation should be coupled with the participation of several high-performing SMEs, universities and others. A Manufacturing Innovation Advisory Group will promote long-term strategic thinking, collective action (and impact), and can highlight best practices for SMEs.

- **Initiate a Collaborative OEM Supplier Upgrade Program**

Most OEMs have their own individual supplier development programs to help suppliers produce efficiently and meet the OEMs' delivery, cost, and quality requirements. However, there is little collaboration across OEMs in the same or different industries when it comes to upgrading the supplier base in the state, even when OEMs share similar suppliers.

Initiatives to upgrade supplier capabilities based on collaboration across OEMs from different industries, like initiatives in Wisconsin by John Deere, Harley-Davidson, etc. to train common suppliers in terms of lean practices, could provide a robust mechanism for leveraging state resources, sharing best practices, and expanding support to SMEs. Such initiatives could focus not only on process and quality improvements but also on technical problem solving and workforce training.

- **Introduce an Advanced Manufacturing SME Innovation Prize**

While several awards for small businesses are already offered in Massachusetts, a state-wide prize for innovative "world-class" advanced manufacturers would not only help set a high bar for SMEs and bring visibility to best practices for SMEs, it would also help change perceptions around advanced manufacturing in the state. The award could be given by a jury comprised of representatives from OEMs, universities, and intermediary organizations who are in a position to identify and evaluate particularly motivated and innovative SMEs.

7.3 Technological and Managerial Support for Innovation in SMEs

- **Provide Technological and Engineering Support**

Thus far, state efforts to support SMEs have largely revolved around workforce training and lean practices. Such practices can lead to greater efficiency and accuracy in terms of quality, cost, and time. However, lean practices are a necessary but not sufficient requirement for success in today's global manufacturing environment. With the rise of new technologies, such as additive manufacturing, programs to support SMEs and build their innovation capacity need to go further. Specifically, support should be expanded to include centers, either existing or yet to be formed, that provide technological and engineering services to SMEs engaged in product and process innovation, like the Connecticut Center for Advanced Technology (CCAT) that offers easy and low-cost access to shared high-end manufacturing machinery for experimentation purposes, such as 3D printing, especially to aviation suppliers.

- **Better Promote and Increase Awareness of Support Services for SMEs**

Although numerous support programs and intermediaries exist in Massachusetts, many SMEs we interviewed were not aware of the portfolio of manufacturing services available in the state. Multiple factors may account for this lack of awareness, but it speaks to the larger challenge of creating an ecosystem that is well connected and where knowledge flows freely. A coordinated communications effort among the various intermediaries that work in this area could help highlight and promote existing support programs and resources within the larger manufacturing ecosystem.

- **Support Executive Education Programs for SMEs**

Advanced manufacturing SMEs are under constant pressure to improve efficiency and innovate. Being "world class" today requires not only a culture and practice of lean, but also sound managerial infrastructure and leadership, combined with a culture and practice of continual product and process innovation.

An executive education program offered by prestigious business and management schools in the state and focused on operations management would help SMEs rise to this challenge and meet a high bar for managerial excellence. Such a program could be offered on a competitive basis and could provide matching funds to support executive education for CEOs and managers at highly motivated SMEs.

7.4 Connections Between Startups and the Innovation Ecosystem

- **Connect SME Capabilities in Early-Stage Scale-Up to the Startup Community**

Many Massachusetts startups, let alone startups outside Massachusetts, are unaware of the deep capabilities that exist within the state to support early-stage prototyping and piloting. Startups currently find manufacturing support through an ad hoc, word-of-mouth process. Efforts by SME trade associations and intermediaries to better communicate these capabilities, together with a more explicit, systematic effort to connect SMEs and startups, is required.

- **Connect Startups with OEMs for Beta Testing and Piloting**

In general, we found it difficult to assess the relative strength or weakness of current links between the Massachusetts startup community and large OEMs in the state. What is clear is that startups are almost always interested in stronger partnerships with potential customers and that more could be done to facilitate such partnerships within the region. Several efforts already exist in particular industries within the state—such as energy and financial services—but more explicit efforts could be geared toward advanced manufacturing-related technologies (e.g., robotics, advanced materials), where development time horizons are longer and where capital requirements during scale-up are higher.

Together these ten system-level recommendations are intended to increase the innovation capacity of the Commonwealth's manufacturing ecosystem through strengthening the links between key nodes within the system. Such steps will build long-term capabilities and institutions for the future that focus on frontier technologies, managerial and operational excellence and connectivity within the ecosystem to ensure Massachusetts' place as a world-class leader in advanced manufacturing.

8 Conclusion

As outlined, Massachusetts has significant assets and expertise in advanced manufacturing that have developed over decades, creating deep capabilities that help to drive innovation in cyber-physical systems in some of the state's leading industry clusters.

But important changes are taking place—within companies and how they are organized for production, in terms of new "game-changing" technologies, and in the global economy as regions and countries work aggressively to increase manufacturing investments and build capabilities.

This changing landscape requires Massachusetts to "up its game" and look to maximize its manufacturing assets in terms of how the key nodes in the state's manufacturing innovation ecosystem are connected, how they collaborate within and across one another, and how innovation is supported and advanced within the system.

To meet the increasingly demanding standards for advanced manufacturing today, we need to set the region on a course of continual upgrading, particularly with respect to small and medium-size manufacturers. The course must also look ahead 5–10 years to ensure we are building the capabilities, the technologies, the workforce and the collaborations that will help fully establish Massachusetts as a world leader in advanced manufacturing.

References

1. acatech (2011) Cyber-physical systems—Innovationsmotor für Mobilität, Gesundheit, Energie und Produktion. In: acatech POSITION, Springer—Band 11
2. Almeida P, Kogut B (1999) Localisation of knowledge and the mobility of engineers in regional networks. Manage Sci 45(7):905–917
3. AMNPO—Advanced Manufacturing National Program Office (2015) NNMI: Snapshot. http://manufacturing.gov/nnmi.html
4. Berger S (2013) Making in America. MIT Press, Cambridge
5. BLS—Bureau of Labor Statistics (2015) International Comparisons of Annual Labor Force Statistics, 1970–2012. http://www.bls.gov/fls/flscomparelf.htm
6. BLS (2015) Quarterly Census of Employment and Wages. Bureau of Labor Statistics. http://data.bls.gov/cgi-bin/dsrv?en
7. Bluestone B, Gartsman A, Walsh D, Eckel R, Huessy J (2012) Staying Power II: a report card on manufacturing in Massachusetts 2012. The Kitty and Michael Dukakis Center for Urban and Regional Policy—School of Public Policy and Urban Affairs—Northeastern University, Boston
8. Bureau of Labor Statistics (2005) Location quotient calculator. http://www.bls.gov/cew/cewlq.htm
9. Chand S, Davis JF (2010) The smart manufacturing revolution. Manuf Exec Leadersh J
10. Communication Promoters Group of the Industry-Science Research Alliance & National Academy of Science and Engineering (2013) Recommendations for implementing the strategic initiative Industrie 4.0—final report of the Industrie 4.0 working group, Frankfurt a.M
11. Edquist C (2005) Systems of innovation approaches—their emergence and characteristics. In: Systems of innovation: Technologies, Institutions, and Organizations. Routledge, Abingdon, pp 1–35
12. EOP—The Executive Office of the President and the U.S. Department of Commerce (2015) Supply chain innovation: strengthening America's small manufacturers, Washington D.C. http://www.esa.doc.gov/sites/default/files/supply_chain_innovation_report.pdf
13. Foss N (1996) Higher-order industrial capabilities and competitive advantage. J Ind Stud 3 (1):1–20
14. Heath S (2003) Embedded systems design. Newnes, Oxford
15. Hounshell D (1984) From the American system to mass production: 1800 to 1932. Johns Hopkins Press, Baltimore
16. IIC—Industrial Internet Consortium (2015) what is the industrial internet? http://www.industrialinternetconsortium.org/about-industrial-internet.htm
17. ITU—International Telecommunication Union (2012) Recommendation ITU-T Y.2060—overview of the Internet of Things. http://www.itu.int/ITU-T/recommendations/rec.aspx?rec=y.2060
18. John Adams Innovation Institute (2011) Building bridges to growth: a roadmap for advanced manufacturing in Massachusetts. Massachusetts Technology Collaborative, Boston
19. Kirkels Y, Duysters G (2010) Brokerage in SME networks. Res Policy 39(3):375–385
20. Kline S, Rosenberg N (1986) An overview of innovation. In: National systems of innovation. Oxford University Press, Oxford
21. Kirner A, Kinkel S, Jaeger A (2009) Innovation paths and the innovation performance of low-technology firms—an empirical analysis of German industry. Res Policy 38:447–458
22. Lerner J (2012) The architecture of innovation: the economics of creative organizations
23. Lester R (2005) Innovation and the competitiveness of local economies, working paper 05-010, MIT Industrial Performance Center, Cambridge
24. Lundvall B, Johnson B, Andersen ES, Dalum B (2002) National systems of production, innovation and competence building. Res Policy 31:213–231
25. MACWIC—Manufacturing Advancement Center Workforce Innovation Collaborative. Employer-led Workforce Solutions. http://www.macwic.org/

26. Maskell M (2001) Towards a knowledge based theory of the geographical cluster. Ind Corp Change 10(4):921–943
27. Massa S, Testa S (2008) Innovation and SMEs: misaligned perspectives and goals among entrepreneurs, academics, and policy makers. Technovation 28:393–407
28. Moore G (1991) Crossing the chasm: marketing and selling high-tech products to mainstream customers
29. Nelson RR, Rosenberg N (1993) Technical innovation and national systems. In: National innovation systems: a comparative analysis. Oxford University Press, Oxford
30. NSB—National Science Board (2012) Diminishing funding and rising expectations: trends and challenges for public research universities, Arlington. https://www.nsf.gov/nsb/publications/2012/nsb1245.pdf
31. OneSource (2015) OneSource global business browser. http://globalbb.onesource.com/web/company/companyfinder.aspx
32. PCAST—President's Council of Advisors on Science and Technology (2011) Report to the president on ensuring American leadership in advanced manufacturing, Washington D.C. https://www.whitehouse.gov/sites/default/files/microsites/ostp/pca
33. PCAST—President's Council of Advisors on Science and Technology (2014) Report to the president: accelerating U.S. advanced manufacturing, Washington D.C. https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/amp20_report_final.pdf
34. Pisano G, Shih W (2011) Producing prosperity. Harvard University Press, Cambridge
35. Reynolds E, Samel H, Lawrence J (2014) Learning by building: complementary assets and the migration of capabilities in U.S. innovative firms. In: Production in the innovation economy. MIT Press, Cambridge, pp 81–107
36. SMLC—Smart Manufacturing Leadership Coalition (2011) Implementing 21st century smart manufacturing—workshop summary report, Washington, D.C.
37. Sturgeon T (2001) How do we define value chains and production networks? IDS Bull 32(3):9–18
38. US Census Bureau (2012) Manufacturing: geographic area series: detailed statistics for the state: 2012. http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ECN_2012_US_31A2&prodType=table
39. USITC—United States International Trade Commission (2010) Small and medium-sized enterprises: U.S. and EU export activities, and barriers and opportunities experienced by U.S. firms. Investigation no. 332–509. http://www.usitc.gov/publications/332/pub4169.pdf
40. Utterback J, Suarez F (1991) Innovation, competition, and industry structure. Res Policy 22(1):1–21
41. Van de Ven AH (1986) Central problems in the management of innovation. Manage Sci 32(5):590–607