

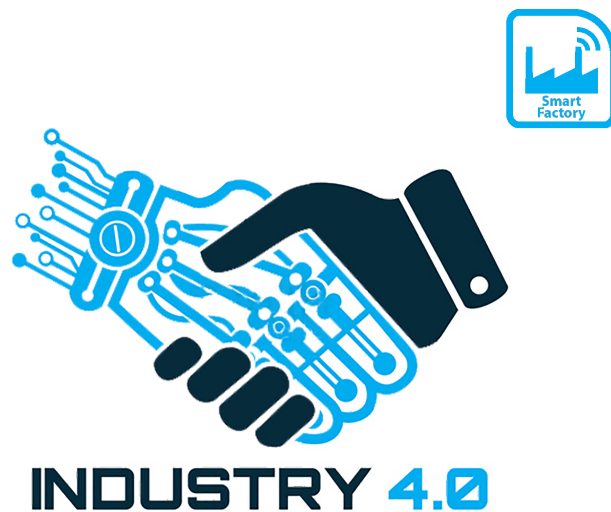
- Lecture Notes -

Industry 4.0 - Smart Factory TWIE24

Industrial Engineering Program

Prof. Dr.-Ing. Stephan Sauter

Q4 2026



Contents

Preface	2
1 Smart Factory - Motivation and Definition of Terms	3
1.1 Goals of the Smart Factory	4
1.2 Internet of Things	5
1.3 Cyber-Physical Systems	7
1.4 Cyber-Physical Production Systems	9
1.5 Smart Factory as a Cyber-Physical (Production) System	10
1.6 Summary	12
2 Development of Automation	13
2.1 Historical Development	14
2.2 The Automation Pyramid	16
2.3 Networked, Decentralized Organization of Production	17
2.4 Five Central Paradigms of Industry 4.0	19
2.5 Future Challenges	20
2.6 Summary	21
3 Technological Foundations and Standards	22
3.1 Identification of Physical Objects	22
3.2 Formal Description Languages and Ontologies	23
3.3 Digital Object Memory	25
3.4 Physical Situation Awareness	26
3.5 (Partially) Autonomous Action and Cooperation	27
3.6 Human-Machine Interaction	28
3.7 Machine-to-Machine Communication	30
3.8 Summary	31
4 Fundamental Concepts of a Smart Factory	32
4.1 Order-Driven Production	32
4.2 Aggregation of Machine and Production Data	33
4.3 Supporting Humans in Production	34
4.4 Intelligent Products and Equipment	36
4.5 Smart Services	37
4.6 Summary	38
5 Reference Architectures	39
5.1 Purpose and Characteristics of Reference Architectures	40
5.2 Overview of Standardization Initiatives	41
5.3 RAMI 4.0 (DIN SPEC 91345)	42
5.4 Summary	44

6	Smart Factory Engineering	45
6.1	Classification of Different Engineering Tools	46
6.2	Virtual Engineering	47
6.3	User-Centered Design	48
6.4	Requirements Engineering	49
6.5	Modeling	50
6.6	Integration of Classic and Smart Components	51
6.7	Summary	53
7	Digital Twin	54
7.1	History	55
7.2	Function	56
7.3	Structure and Optimization	57
7.4	Key Characteristics	59
7.5	Challenges and Risks	59
7.6	Opportunities	60
7.7	Examples in Practice	61
7.8	The Future of the Digital Twin	62
7.9	Summary	64
8	Security Risks in the Smart Factory	65
8.1	IT Security	66
8.2	Data Protection	67
8.3	Organizational Risks	67
8.4	Economic Risks	68
8.5	Industrial IT Security According to IEC 62443 and ISO 27001	68
8.6	Recommendations for Action by the BMWi	71
8.7	VDMA Guideline According to IEC 62443	71
8.8	Summary	72
9	Sustainability in Production	74
9.1	Sustainability	74
9.2	Sustainability at the Company Level	75
9.3	Types of Sustainability	76
9.4	Opportunities and Benefits of Sustainable Product Policy	77
9.5	Goals of Sustainable Production	77
9.6	Concepts and Methods	78
9.6.1	Sustainable Supply Chain Management	79
9.6.2	Circular Economy	80
9.6.3	Material Flow Management and Analyses	81
9.6.4	Product Carbon Footprint	83
9.7	Siemens as a Pioneer of Sustainable Production	83
9.8	Summary	86
10	Smart Logistics	87
10.1	Future Challenges and Trends	88
10.2	Predictive Analytics in Logistics	88
10.3	Prescriptive Analytics in Logistics	89
10.4	Extralogistics	91
10.5	Intralogistics	95
10.6	Summary	101
11	IO-Link	102
11.1	Automation Systems and Their Requirements	103

11.2 Consortium and Brand	104
11.3 Who Manufactures IO-Link Devices and Masters?	105
11.4 How Does IO-Link Work?	105
11.5 IODD	107
11.6 Advantages of IO-Link	108
11.7 Summary	110
12 Predictive Maintenance	111
12.1 Overview of Maintenance Strategies	112
12.2 Definition and Objectives of Predictive Maintenance and Smart Factory	113
12.3 Technical Foundations	114
12.4 Predictive Maintenance in the Smart Factory	115
12.5 Forecasting Concepts and Remaining Useful Life	117
12.6 Summary	119
13 Machine Learning in Industry	120
13.1 What is Machine Learning?	121
13.2 The Learning Process	121
13.3 Learning Types	122
13.4 Models in Machine Learning	124
13.5 Data as a Foundation	125
13.6 Multimodal Data and Information Fusion	126
13.7 Applications in Industry	127
13.8 Limitations and Challenges	128
13.9 Summary	129
14 Semiconductor Industry	130
14.1 Relevance of Semiconductors in Industry 4.0	131
14.2 Fundamentals of Semiconductor Technology	132
14.3 Lithography as the Key Process in Semiconductor Manufacturing	134
14.4 Structuring and Doping of Semiconductors	135
14.5 Semiconductor Technology in the Smart Factory	136
14.6 Supply Chains, Production, and Logistics of Semiconductors	138
14.7 Challenges and Trends	139
14.8 Summary	140
15 Robotics in the Smart Factory	141
15.1 Historical Development and Terminology	141
15.2 Fundamentals of Robotics	142
15.3 Applications of Robots in Industry	143
15.4 Types of Robots in Industry	143
15.5 Humanoid Robots	145
15.6 Internet of Robotic Things	147
15.7 Lot Size 1 in Industry 4.0	150
15.8 Summary	150
Bibliography	153

Preface

This script is intended to help the listener follow the lecture through independent study, alongside the online session. The exam will exclusively contain questions on selected chapters of the script. The literature cited is intended to support a deeper understanding. Online knowledge quizzes offer a parallel opportunity for self-assessment.

Course content:

After attending this lecture, students should know the term Smart Factory and be able to place it within the context of Industry 4.0. They should be able to trace the development of automation up to the fully autonomous, decentrally organized production plant.

The fundamental technologies and standards used for the design and operation of a Smart Factory are introduced. Students learn the essential concepts of a Smart Factory and are able to identify and distinguish the individual elements using various reference architectures.

The particular engineering challenges are examined, specific security risks of digitalized and networked production plants are presented, and concrete recommendations for action are assigned to each.

Core Literature:

- Westkämper, E., Spath, D. et al (2013): Digitale Produktion. Springer, Stuttgart [1]
- Bauernhansl (2014): Industrie 4.0 in Produktion, Automatisierung und Logistik, Springer [2]
- Handbuch Industrie 4.0 Bd. 1-4 (2023), Springer Nature [3]



Links:

- <https://www.plattform-i40.de>
- www.freiwilligschlauwerden.de

CHAPTER 1

Smart Factory - Motivation and Definition of Terms

Learning Objectives

After working through this chapter, you will know ...

- ... what is meant by Smart Factory.
- ... what Industry 4.0 is all about.
- ... what »digital transformation« means.
- ... which opportunities and challenges the Smart Factory offers.
- ... how the Smart Factory goal can be achieved.

Introduction

The term smart production or »Smart Factory« is not clearly defined. As part of the (likewise not clearly defined) research initiative **Industry 4.0**¹ and the comprehensive conversion of production to digital technologies, the Smart Factory occupies a central position for the future development of industrial production.

In order to survive in an increasingly complex market environment, companies must preserve their competitiveness and adopt measures that will sustainably change production and corporate organization.

Examples of once market-dominating large companies that missed technological disruptions are Kodak (photography industry, digitalization), Yahoo (search engine technology), or Nokia (mobile phones).

The new possibilities offer medium-sized, small, or even micro-enterprises (SMEs) opportunities for entirely new business fields, in which they can even achieve a market-dominating position (e.g. the Viv app for speech recognition, today Bixby Voice)



¹ **Industry 4.0** refers to a future-oriented project for the comprehensive digitalization of industrial production in order to better equip it for the future.

1.1 Goals of the Smart Factory

The Smart Factory (German »intelligente Fabrik«) is at the center of the future initiative **Industry 4.0** launched by the German federal government in 2012. The goal is to provide a production environment that, through the massive use of the internet and the digitalization of technical processes and business processes, ideally organizes itself.

The Smart Factory consists of manufacturing facilities and logistics systems whose core components are **Cyber-Physical Systems (CPS)**, which are intelligently networked with one another. Ideally, humans no longer need to intervene in the actual production process.

Compared to earlier production methods, the Smart Factory offers a wide range of advantages and possibilities:

Three main goals and opportunities of the Smart Factory include, among others:

1. Individualization:

- Taking short-term customer requirements into account during planning, design, and production;
- Production of individualized products (batch size 1) at prices comparable to mass-produced products.

2. Flexibilization:

- greater flexibility in production (e.g. through digital engineering, 3D printing for prototypes)
- rapid implementation of innovations
- shorter time-to-market for new products
- consumption-driven supply of production
- rapid adaptation to new or changed product requirements

3. Productivity Increase:

- lower production costs (e.g. through optimization of resource and energy efficiency)
- predictive maintenance (predictive maintenance & service)
- shorter production times
- automated, efficient, and transparent ordering processes [4]

In order to achieve these goals, however, numerous challenges must be mastered. As the requirements of the market environment increase, and as the performance capability and diversity of production systems grow, delivery capability and availability must increase accordingly.

The advancing »**digital transformation**«¹, i.e. the introduction of digital technologies in all areas of production, is intended to meet these increased requirements.

¹ **Digital transformation**, or digital change, refers to the transition from analog processes to digitalized procedures, primarily in industry.

1.2 Internet of Things

In the **Internet of Things (IoT)** (German: »Internet der Dinge«), objects can communicate with one another and give or receive commands. The Internet of Things thus makes it possible to carry out tasks and applications automatically, without external control.

Since there is no universally valid definition of the Internet of Things, descriptions of IoT can differ considerably from one another. Depending on the field of application and the technology used by the network, different interpretations are possible. In general, the term »Internet of Things« is used for the networking, via the internet, of everyday objects or of production facilities in manufacturing industry.

In general, a distinction can be made between the areas of »private use« and »industrial applications«. In contrast to private use, where the comfort and entertainment provided by intelligently networked everyday objects are paramount, in the industrial domain safety and revenue growth are essential. Applications of IoT include, for example, intelligent building automation, or devices that contact the user via the internet when certain events occur.

All devices connected to one another in the network are assigned a unique identity in the network (address). Through embedded electronic intelligence, participating devices can make contact with one another and carry out tasks fully automatically. This is why these objects are also called **Smart Devices**.

The Technical Foundations of the Internet of Things

The internet continues to grow due to the strongly increasing number of participating devices. The reason for this lies in ever cheaper, smaller, and more powerful microcomputers. With little effort, objects can thus be upgraded with intelligent electronics and connected to the internet with their own address, either by wire or even wirelessly.

A basic prerequisite for managing the exponentially growing demand for unique addresses for each individual object is Internet Protocol Version 6 (**IPv6**), introduced in 1999, which has replaced Version 4 (**IPv4**).

IPv6 uses 128-bit addresses (IPv4: 32 bits, 4.3 billion addresses), which results in an inconceivably large number of addresses. In total there are 2^{128} possible addresses, which corresponds to a 3 followed by 38 zeros.

	Internet Protocol version 4 (IPv4)	Internet Protocol version 6 (IPv6)
Deployed	1981	1999
Address Size	32-bit number	128-bit number
Address Format	Dotted Decimal Notation: 192.149.252.76	Hexadecimal Notation: 3FFE:F200:0234:AB00: 0123:4567:8901:ABCD
Prefix Notation	192.149.0.0/24	3FFE:F200:0234::/48
Number of Addresses	$2^{32} = \sim 4,294,967,296$	$2^{128} = \sim 340,282,366,920,938,463,463,374,607,431,768,211,456$

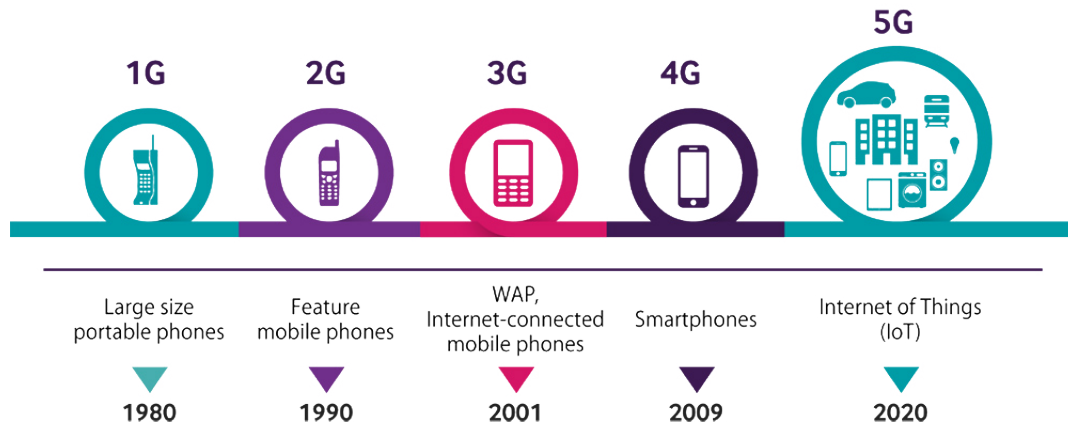
Comparison of IPv4 and IPv6

The general concept is the same for IPv4 and IPv6: every device (computer, tablet, smartphone, router, etc.) has a routing table, based on which it decides how an IP packet should be forwarded.

IPv6 offers major advantages through increased security, as it includes an integrated security function (IPsec).

Connected devices send or receive their data and commands via these protocols. In addition to these protocols, wireless communication technologies also exist, such as WLAN, Bluetooth, or mobile communication standards such as UMTS (3G) and LTE (4G).

The 5G standard, which replaced the LTE network in 2020, is about 10 times faster and optimized for video and streaming services and real-time applications, making it predestined for Industry 4.0 applications [5].



According to the German Federal Ministry of Education and Research, the 6G network, the sixth generation of mobile communications, is expected to succeed the 5G network in Germany from around 2030, with transmission rates up to 100 times higher, reaching up to 1000 GBit/s.

	INTRODUCTION YEAR	TOP DOWNLOAD SPEED	TIME TO DOWNLOAD A 3GB MOVIE
1G	1980	2 Kbps	6 days
2G	1990	100 Kbps	2.5 hours
3G	2001	8 Mbps	2 minutes
4G	2009	150 Mbps	20 seconds
5G	2020	10 Gbps	300 milliseconds
6G	2030	1 Tbps	3 milliseconds

According to chip manufacturer Qualcomm, AI applications in particular will benefit from 6G. As an example, Qualcomm cites autonomous driving, where large amounts of data are exchanged in real time between vehicles and infrastructure, so high speeds are advantageous. The Fraunhofer Institute also speaks of advances in medicine, since sensors could evaluate health data within the shortest possible time.

Applications of the Internet of Things in the Smart Factory

For industrial enterprises, IoT is the optimal means of reducing costs, increasing efficiency, and generating additional revenue through new business models. While manufacturers have long had access to the data collected in production, the data in manufacturing's own closed software silos was inaccessible. This limits the ability to use this data for decision-making.

IoT makes it possible to collect and manage large volumes of manufacturing data in the cloud, not only within a single factory but across multiple production sites. When this data is combined with analytics and evaluated, companies gain much better insight into their processes,

enabling them to optimize operations, reduce quality defects, and implement preventive maintenance.

1.3 Cyber-Physical Systems

Information technology is experiencing rapid progress due to the ever-faster increase in the performance capability of microprocessors. In complex systems that are integrated into buildings, products, and production facilities, increasingly intelligent, software-intensive solutions are being applied.

These so-called embedded systems (**embedded systems**)¹ are able to communicate with one another through mutual networking. This makes it possible to intelligently link data and services via the internet. The physical world is connected with the software world, giving rise to smart solutions in which humans only monitor and control by interacting with software platforms.

The core components that enable these solutions are referred to as Cyber-Physical Systems (CPS). CPS is a term originating in the USA that has become an omnipresent term in future technical and industrial trends and challenges.

Through the use of these Cyber-Physical Systems, the internet becomes „**real world aware**“ (**RWA**)² and is able to act directly based on physical reality. Through the possibility of interaction, physical reality and virtual space (**cyberspace**)³ grow together into a single unit.

Cyber-Physical Systems are the foundation of the Smart Factory and of the intelligent networking of machines and products. Because products are equipped with CPS, they can provide all relevant information about the intelligent factory. Based on this data, the control of the individual production steps can then be carried out in a fully automated, end-to-end manner.

CPS comprise embedded systems, production, logistics, engineering, coordination, and management processes, as well as internet services [6].

$$\text{CPS} = \text{Embedded System} + \text{Intelligent Networking}$$

A Cyber-Physical System consists of one part in physical reality and one part in digital, or virtual, space:

- In the physical world, the CPS consists of sensors (measuring), actuators (acting), user interfaces, and the connecting digital communication technology (communicating).
- At the logical level, intelligent functions handle the acquisition, transport, processing, and provision of information (computing + communicating).

¹ An **embedded system** is an electronic computer that is integrated (embedded) into a technical context.

² **RWA, Real World Awareness** means a timely representation of the real world in IT systems.

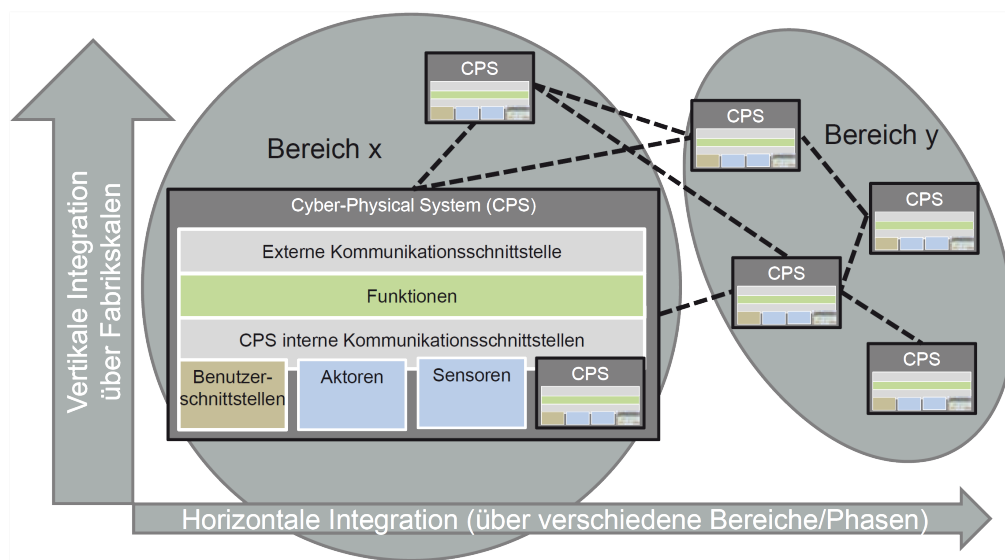


Different Levels of Cyber-Physical Systems

New solutions become possible through the consistent application of IP communication technologies, software services, and global, open standards in mechatronic systems. Networking enables interaction in two directions, which is the prerequisite for both receiving and sending commands.

The type of networking within a Smart Factory can be divided into two areas:

- Vertical networking between the different levels of the factory, ranging from the technical process within a production facility, workstations, and manufacturing cells up to the entire production network.
- Horizontal networking between different CPS at the same level or organizational units, such as between two manufacturing processes or between production and product development.



Structure of a Cyber-Physical System

There are various characteristics of CPS that can be classified according to their degree of spatial distribution:

- spatially contained within a single housing, with all functions of a CPS integrated. Example: „Smart Device“
- the functions are distributed across many IT systems around the world (highly distributed application). Example: networked factories of a company in different countries.

1.4 Cyber-Physical Production Systems

The use of CPS across entire production facilities is referred to as **Cyber Physical Production Systems (CPPS)**¹.

Translated, CPPS means: a system of intelligent CPS products and CPS resources/equipment. This refers to the consistent integration of CPS into production facilities, warehousing systems, and means of production, which constantly exchange and process information with one another. This gives rise to intelligently networked factories and entirely new value creation networks.

These value creation networks are based on the shared capability of their components to communicate with one another. This is made possible by components such as sensors integrated into the systems of the production facilities. For example, **RFID chips**² can be used with these facilities to exchange information about equipment and the transport system.

Networks become possible that use this data in real time to control and organize themselves.

”Manufacturing thus becomes fully transparent end-to-end” [4]. Embedded CPS unfold their full performance capability especially in connection with the internet. Cyber-Physical Production Systems are characterized by:

- real-time acquisition of physical data via sensors;
- use of globally available data and services;
- storage and analysis of data;
- networking via digital communication (local/global);
- influence on real objects via actuators;
- use of human-machine interfaces [7]

CPPS shape the future of production and logistics through:

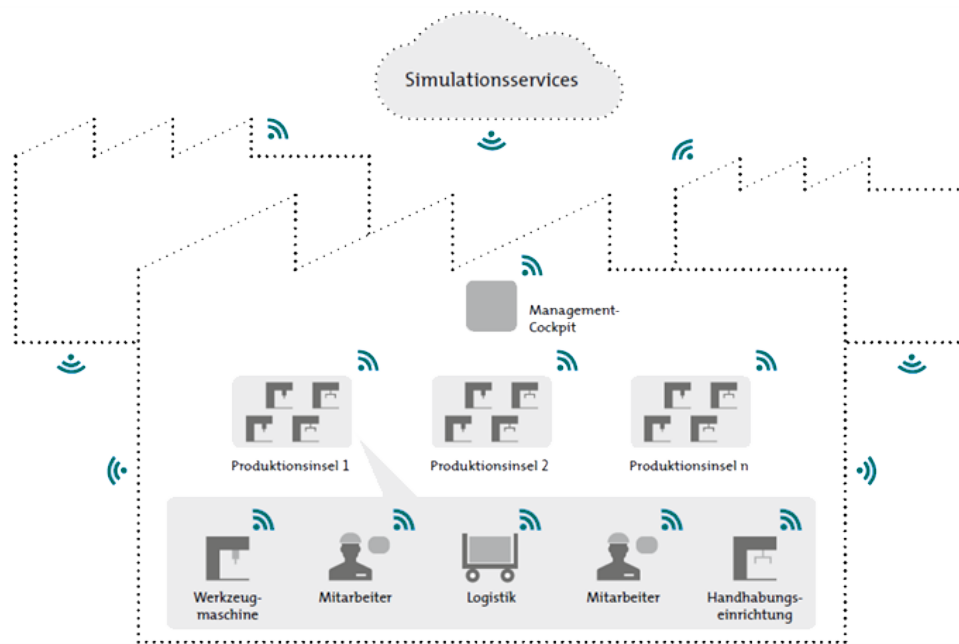
- smart products, as well as
- smart production or operating resources.

CPPS enable decentralized, flexible production and logistics control. The increased use of centrally available data from sensors and local control loops thus leads to situation-based decision alternatives, e.g. using virtual models that represent the physical world.

All areas of the factory are thus networked with one another and can communicate in both directions, providing decision-making options (e.g. for the management cockpit) via simulation services.

¹ **CPPS** is a system of intelligent CPS products and CPS resources/equipment.

² **RFID chip** stands for Radio Frequency Identification Chip.

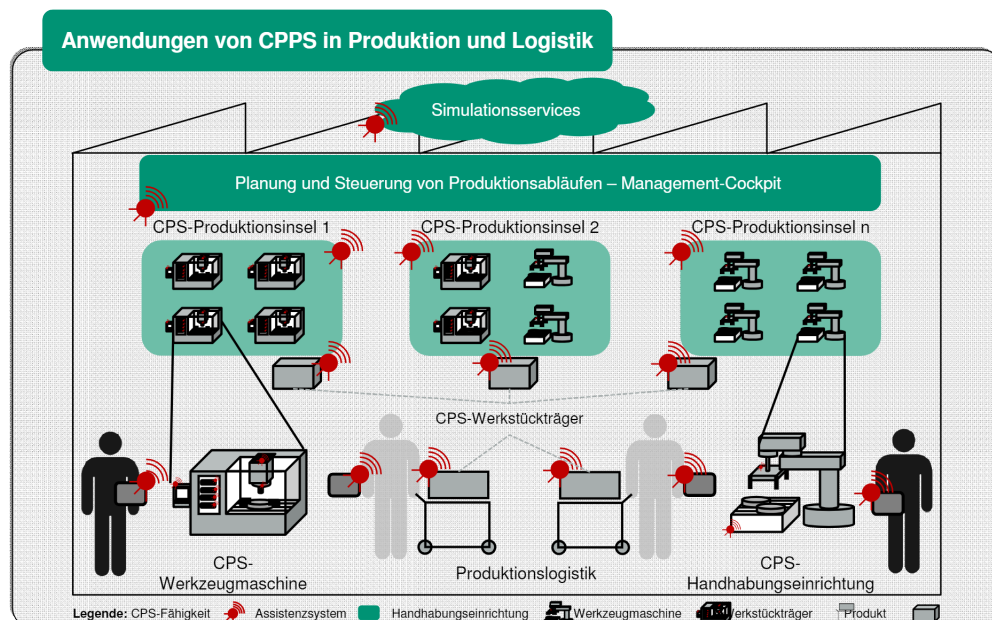


Cyber-Physical Production System (CPPS)

1.5 Smart Factory as a Cyber-Physical (Production) System

The so-called Smart Factory emerges through the autonomous and decentralized networking of individual Cyber-Physical Systems. In interaction with humans, these networks can solve problems independently.

A key characteristic here is that information from the factory is available in real time, thereby uniting the physical with the virtual world.



CPPS, Cyber-Physical (Production) System

The foundation of an intelligent factory is formed by so-called active workpiece carriers, on which one or more workpieces or products are mounted so they can be transported by a conveying

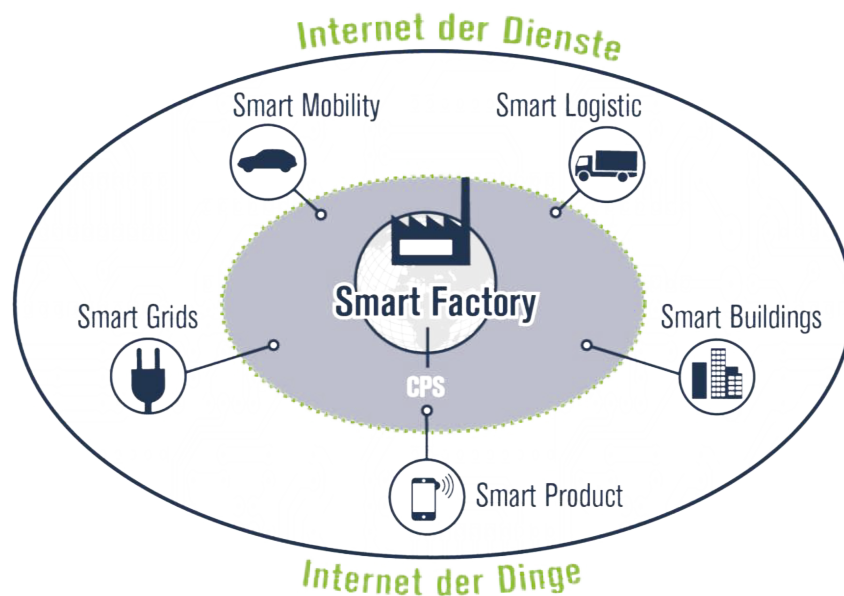
mechanism along a production line through, in some cases, multiple production stations. These workpiece carriers can be understood as CPPS.

The advantages of such active workpiece carriers acting as CPPS are:

- greater flexibility;
- the workpiece carrier takes over functionality from the production line;
- sensors, logic, and actuators are built up modularly;
- standardized interfaces enable traceability;
- continuous data acquisition throughout the entire production process;
- more efficient production processes;
- throughput times are reduced through active CPS use [7]

Through CPS and the use of intelligent methods and processes, the main focus is thereby directed toward production. In addition, intelligent products actively support the production process.

At the interfaces, the Smart Factory thus becomes the subject of a comprehensive intelligent infrastructure.



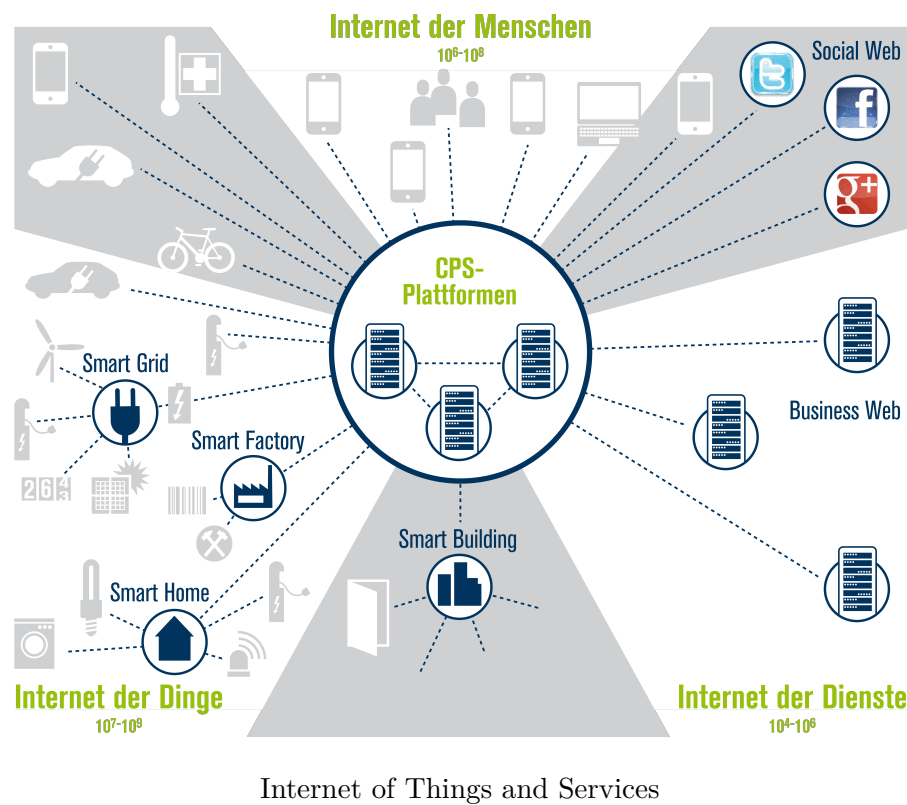
Cyber-Physical Systems Enable the Intelligent Factory

In the future, CPS platforms will emerge that network people, objects, and systems far beyond a single company through applications and services.

The CPS platforms form the basis for linking the various networks with one another: the internet of people with the internet of things and the internet of services.

„With the increasing use of the internet of things and services in production, value chains will change in combination with smart production, smart logistics, smart grids, and smart products, and new business models will become established.

The future project Industry 4.0 will unlock existing technological and economic potential. Industry 4.0 opens up new business perspectives and new innovative social infrastructures for work. [4]



1.6 Summary

The Smart Factory is a sub-area of the Industry 4.0 initiative and ideally stands at the end point of the ever-advancing digitalization of production. The fundamental goals are to preserve the competitiveness of companies through cost reduction, entry into new markets, and the associated creation of new products and services. Achieving these goals requires numerous changes, the foundation of which is the comprehensive digitalization of production and control.

Digital technologies such as the Internet of Things (IoT), Cyber-Physical Systems (CPS), and their integration into the production process are the decisive factors here for ideally creating a completely autonomous factory.

In the future, the Smart Factory should be able to control itself, in real time, as a Cyber-Physical (Production) System (CPPS). To this end, the Internet of Things must, for example, be able to address far more devices and products. Moreover, mobile networks must also be expanded to handle the growing number of participants, volumes of data, and speed (real time).

It will likewise be necessary for the various networks (internet of things, internet of services, social networks) to be linked together on so-called CPS platforms.

CHAPTER 2

Development of Automation

Learning Objectives

After working through this chapter, you will know...

- ... how automation has developed historically.
- ... what the automation pyramid is.
- ... how future production will be organized.
- ... how the transition to the Smart Factory can succeed.
- ... which challenges and goals need to be mastered.
- ... how these goals can be achieved.

Introduction

The economic efficiency of production processes is often impaired by the need for manual preparation of digital production data and the manual adaptation of data to individual component geometries in component-specific production. The networking of production systems such as manufacturing or measuring machines and of planning and simulation systems leads to a significant acceleration of preparation and adaptation processes.

The planned objectives are to increase productivity and improve the flexibility of manufacturing through digitalization and networking in production.



Figure 2.1: Automation in vehicle production

2.1 Historical Development

A process or facility is referred to as automatic if it runs or operates under defined conditions without human intervention. Automation therefore merely describes whether a process takes place independently and without the active involvement of living beings. Industry 4.0 is only spoken of once cyber-physical systems are operated in an interconnected manner. Industry 4.0 is thus understood as the fourth stage of automation, in which intelligent and (partially) autonomous entities interact.

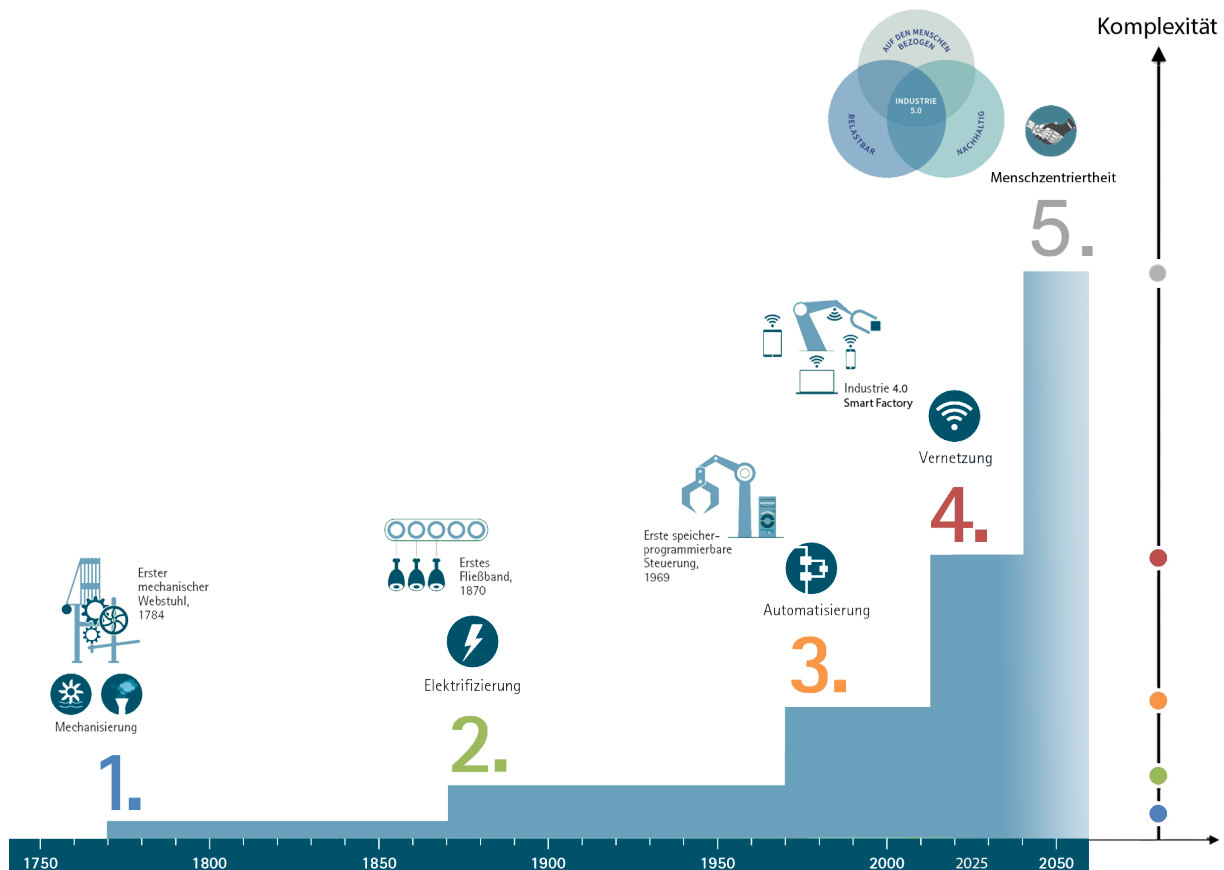


Figure 2.2: Industrial Revolution – Industry 1.0 to 5.0

- Until the middle of the 18th century, all production was accomplished through the muscular power of humans and animals. The invention of the steam engine ushered in the **first industrial revolution** from this point onward.

During this period, muscular power was replaced by water and steam power, thereby driving forward mechanization. Products could suddenly be manufactured faster and in greater quantities, and this era lasted roughly 100 years, up to the middle of the 19th century.

- The **second industrial revolution** superseded this, as mass production was introduced with the help of electrical energy and assembly lines. This period, too, lasted roughly another 100 years.
- In the **third revolution, also called the digital revolution**, information technologies began to be used from around 1970 onward to automate production. Work steps that had previously only been possible to carry out by hand could now be taken over by machines. After roughly 50 years, we still find ourselves in this era today – but on the threshold of the fourth (and fifth) industrial revolution.
- The term **Industry 4.0** was published for the first time in 2011 at the Hannover Messe. On April 8, 2013, the Industrie 4.0 Working Group, chaired by Siegfried Dais (Robert Bosch

GmbH) and Henning Kagermann (acatech), handed over its final report at the Hannover Messe. Since then, the Plattform Industrie 4.0 has taken up its work. Here, the three industry associations Bitkom, VDMA, and ZVEI coordinate future activities.

	Gestern - Klassisch (3. industrielle Revolution)	Heute - Industrie 4.0 (4. industrielle Revolution)
Produktionssteuerung	Zentral	Dezentral (CPS, Cloud)
Steuerungstechnik	Monolith	Offener Standard im Netz (Cloud)
Datenverarbeitung und -bereitstellung	Zeitversetzt	Echtzeit-Informationen
Nutzungsrechte (Softwarekosten)	Lizenz(-kosten)	Pay-per-use
Softwareverwaltung	Software-Suite	Cloud-Apps (SaaS)

Figure 2.3: Industrial Revolution – yesterday and today

In many industries and SMEs, we currently still find ourselves in transition from classic industry toward Industry 4.0. As can be observed in virtually all companies and processes, decentralization, networking, and digitalization represent the decisive changes here. An exact point in time at which the transition to Industry 4.0 will be finally completed cannot be named, since advancing digitalization continually creates new possibilities for networking, decentralization, and data processing (AI).

- **Industry 5.0**, unlike the by-now-established Industry 4.0 guiding concept, is still at an early, continuously evolving stage. It can be assumed that in the coming years, increasingly concrete implementation aids – standards, maturity models, funding programs, and collections of best practices – will emerge.

The three "new" aspects/dimensions of **sustainability**, **resilience**, and **human-centricity**, which represent the core elements of the step toward Industry 5.0, considerably increase complexity further. Whether **AI** as an enabling technology for a fifth industrial revolution possesses a sufficiently disruptive character to justify the increment in numbering is an ongoing discussion among experts (<https://www.acatech.de/die-vierte-industrielle-revolution-und-industrie-5-0-eine-kritik/>, Noha M. Galal, 2025, A Review of the Industry 4.0 to 5.0 Transition).

The transformation toward Industry 5.0 is, however, carried by the trend of designing further developments in a completely **technology- and platform-independent** manner.

A closer interlinking of Industry 5.0 with adjacent regulatory developments is also foreseeable, in particular with the European legal framework for artificial intelligence (the AI Act) as well as with reporting obligations regarding sustainability and supply chain responsibility. For manufacturing companies, this means that human-centered, resilient, and sustainable design is increasingly likely to become not merely voluntary commitment but regulatorily expected practice.

For corporate practice, this yields the recommendation to understand the transition from Industry 4.0 to Industry 5.0 not as an abrupt break but as an **evolutionary extension** of existing digitalization strategies: existing Industry 4.0 infrastructure – networking, data analytics, automation – already forms the technical basis on which human-centricity, resilience, and sustainability can be specifically built.

2.2 The Automation Pyramid

Automation can take place at every level of a company. This is illustrated by the automation pyramid, which depicts communication methods within a company and is shown in the figure below. The following levels within a company are distinguished [8]:

- **Enterprise management level:** It makes strategic decisions that affect the company as a whole. Tasks such as market analysis, corporate management, and strategic personnel, investment, and production planning are carried out here. At this level, systems such as Enterprise Resource Planning (ERP) or SAP products are used.
- **Plant management level:** It deals with the planning and execution of concrete orders, deadline monitoring, and cost analyses. Software solutions such as the Manufacturing Execution System (MES) are used here.
- **Production control level:** This level manages short-term production planning such as the scheduling of machines, systems, and personnel. Software solutions for this are so-called Supervisory Control and Data Acquisition (SCADA) systems.
- **Process control level:** depending on plant size, the process control level can be further subdivided into the plant, group, and unit control levels. Since the processes take place within a manufacturing area (cell), this level is also called the cell level from the perspective of the manufacturing facilities.
- **Field level:** It comprises individual actuators, sensors, and display devices. At the field level, data is processed and corresponding responses are triggered. It connects the process with the higher-level controllers. For the connection to the higher-level systems, it is important that data volumes are transmitted with short response times.

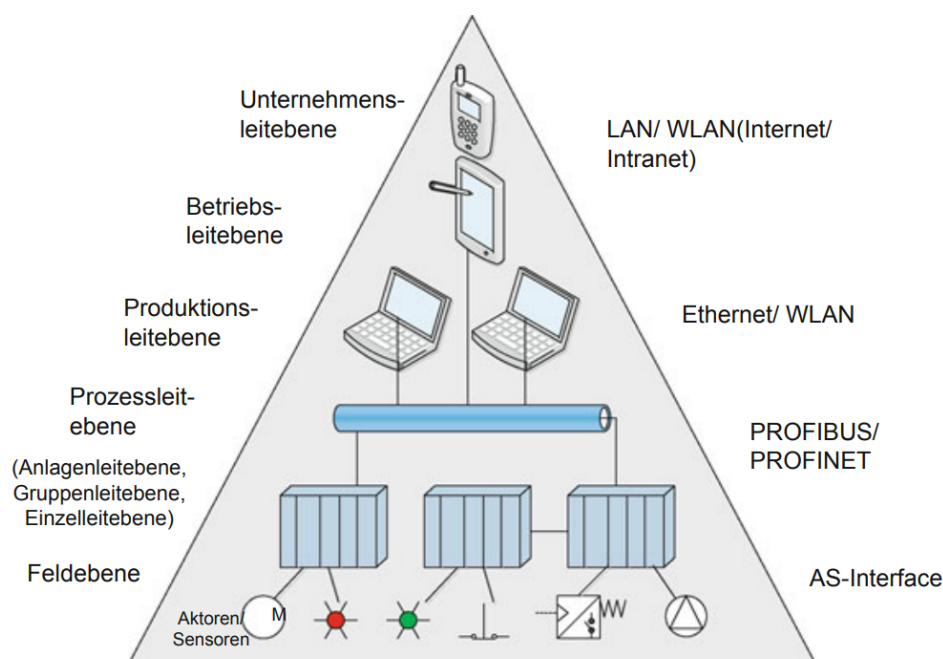


Figure 2.4: The automation pyramid

Accordingly, the level of detail in the automation pyramid increases from top to bottom, while the level of abstraction decreases. Depending on the hierarchy level within the automation pyramid, different communication tools are used. At the upper levels, complex computer systems are found in networks of large scale with many participants. Processing times are not of decisive importance here, since the data calculated at this level is used for long-term planning. At the lower levels, by contrast, response times are of high importance, since real-time data is processed here and immediate reactions are planned.

While both LAN and WLAN technologies are used at the enterprise management level and the plant management level, Ethernet/WLAN-type communication techniques are used at the production control level. The process control level uses PROFIBUS/PROFINET, and the field level uses AS-Interface technologies. It can accordingly be concluded that the hierarchy of the automation pyramid is also reflected in the choice of technologies used for the exchange of information.

2.3 Networked, Decentralized Organization of Production

In the previous deterministic, centralized planning of production, a shared and optimized production plan was assumed. Until now, it was common practice to centrally determine and temporally coordinate production tasks such as the procurement of resources. The results were often only moderately satisfactory. With the increasing complexity of the various processes in the individual areas, new **decentralized approaches** are necessary.

Decentralized control systems are designed to optimize tasks and priorities locally. Compared to centralized control, attention is confined to a limited and manageable scope. It is therefore much easier to respond flexibly to short-term changing requirements through local optimizations. In decentralized planning, the executing level can furthermore be directly involved, meaning short-term goals (e.g., releasing production capacity) can be implemented much more efficiently.

Decentralized and mostly short-term control by intelligent and networked technologies presupposes recurring situations, so that they can be automatically recognized by the software. Decentralized control also makes it possible, through delegation to the execution level, to implement standardization more easily, since "only" the interfaces need to be unified. Through decentralized control measures, routine decisions are standardized by intelligent assistants.

Decentralized control provides answers to the rapidly changing, ever more complex environment of companies:

Reduction of Planning Capacity

The decisive advantage of decentralized planning and control lies in the **reduction of planning capacity**. Since external complexity continues to rise due to, e.g., more flexible customer requirements, internal complexity in production or within the company must also increase in order to be able to serve these needs.

Increasing External Complexity

The external market perspective is referred to as external complexity and includes changes in the market (demographic change, digitalization, etc.), increasing flexibility (variant diversity...), and the actual products themselves (price, functionality, variety).

Increasing Internal Complexity

From the company's perspective (internal complexity), the increased external requirements must likewise be answered with increased complexity. This can only be achieved if the planning and control mechanisms are outsourced to the individual areas of the company, i.e., anchored in a decentralized manner.

Since not all control mechanisms can be immediately converted to a decentralized model (e.g., firmly anchored IT systems), a balanced, effective, and efficient middle ground must be found. Through decentralization and networking, high currency, and continuous availability of information, it is thus "only" a small step to autonomous and self-organizing production.

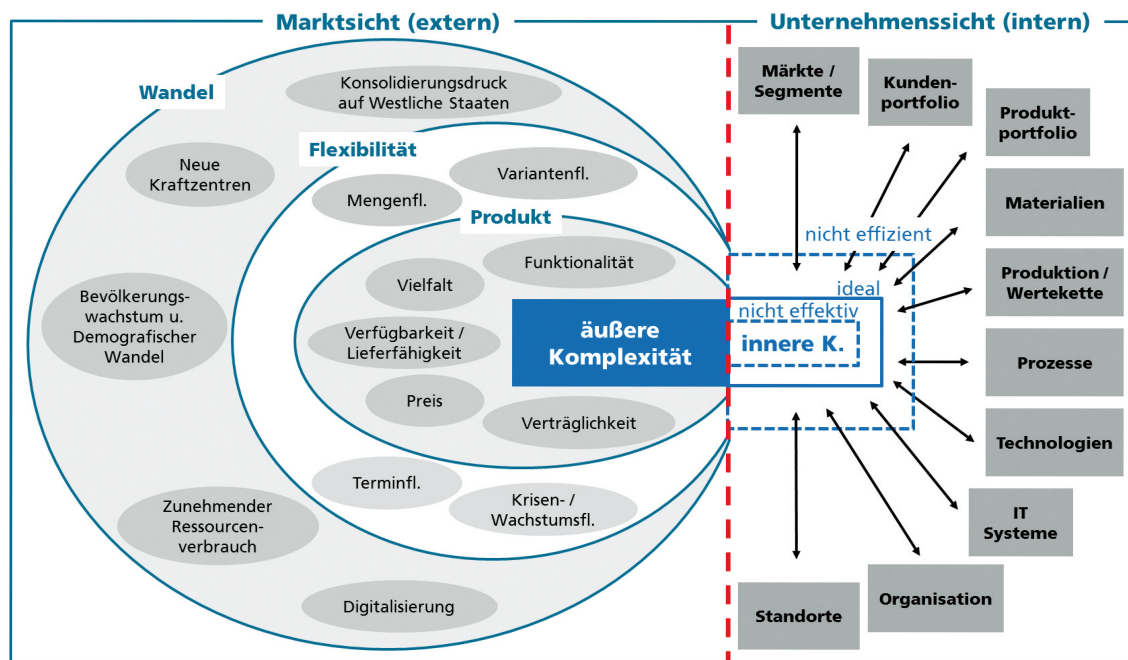


Figure 2.5: Internal complexity must correspond to external complexity. The challenge for the company is to achieve the respectively optimal level of complexity

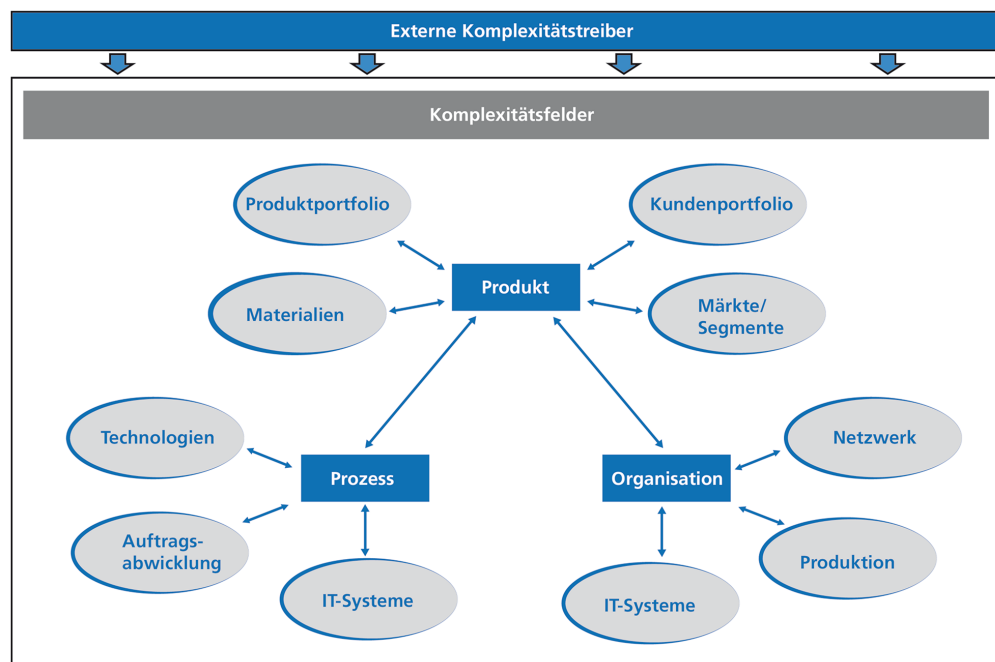


Figure 2.6: Decentralized control and the resulting outsourcing of complexity to the individual work areas

2.4 Five Central Paradigms of Industry 4.0

The term "Industry 4.0" is often related merely to the use of novel technologies. Even though the individual technologies play an important role in the shift toward Industry 4.0, they are often not genuinely novel. For example, intelligent systems and broadband internet access play a major role in Industry 4.0, even though these are not fundamentally technological novelties. Rather, the true novelty and innovation of Industry 4.0 lies in the integration of all these systems. The following five paradigms describe the changed ways of thinking compared to traditional production approaches.

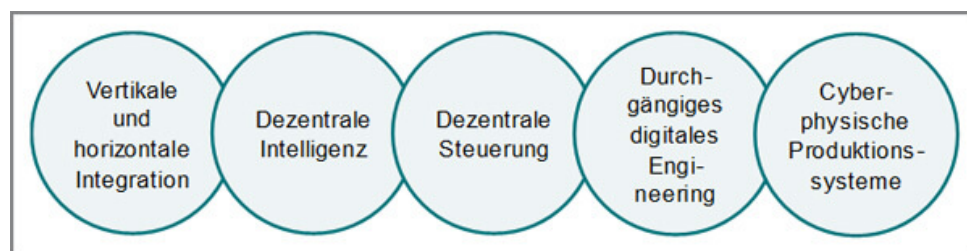


Figure 2.7: Novel paradigms of industrial manufacturing

- **Vertical and Horizontal Integration:** Efficient data collection can only be ensured through vertical integration of the relevant systems. This means that all systems must be hierarchically organized and connected through suitable interfaces. Only in this way can an orderly data flow be ensured and information exchanged efficiently. This is described by the automation pyramid. Horizontal integration describes the incorporation of customer, supplier, and other external systems into a company's own vertical system landscape. This enables the exchange of data both within and outside the company in order to maximize material flow and thus efficiency.
- **Decentralized Intelligence:** Decentralized intelligence is a prerequisite for decentralized control and describes the ability of means of production and production facilities to

individually and location-independently pass on relevant information to a decentralized control system. Technologies such as the Internet of Things and Services are used for this purpose, since these systems can flexibly forward and process data.

- **Decentralized Control:** In decentralized control there are no central decision-making mechanisms and rigid boundaries. Decentralized control enables dynamic adaptation and scaling of the required computing power. Access to production control no longer requires direct or location-bound access to a central computing unit here. This now takes place merely via an internet-enabled end device.
- **Consistent Digital Engineering:** Consistent digital engineering refers to the digital representation of a complete physical production process. The physical and virtual worlds mesh seamlessly, and all processes can be visualized and edited in real time. This allows real-world applications to be digitally simulated and analyzed.
- **Cyber-Physical Production Systems:** A Cyber-Physical Production System (CPPS) describes a production setup that fulfills the four previously mentioned paradigms. When all these conditions are met, this is also referred to as a cooperating overall system.

2.5 Future Challenges

Value creation in general will in the future take place within complex process landscapes. To remain competitive in this environment, companies must keep pace at all levels and digitalize and automate their value creation processes. In doing so, the advantages that result from this should not obscure the fact that considerable risks are also associated with it. Answers to these challenges must be found already now.

Production companies must find answers to the following challenges:

- **Loss of control over one's own production processes:** The scope for entrepreneurial decision-making is reduced by the automation and networking of production and business processes. This can lead to certain processes no longer being able to be intervened in.
- **Security and digital attacks:** Through networking via the public network, companies are exposed to digital attacks that can spread across the entire value chain. To prevent manipulation and espionage, IT systems must be appropriately secured with security architectures.
- **Dependence on large technology providers:** The scope for design and decision-making can be extremely restricted by the use of platforms provided by a few large technology companies.
- **Production downtime:** Due to the networking of all production facilities, even small incidents can lead to the complete failure of the production line.
- **Investments:** The introduction of new technologies requires investments at all levels. The overriding principle must be economic profitability, so that, for example, investment is made only in market-ready technologies.
- **Automation threatens jobs:** Advancing digitalization and automation inevitably leads to employee activities being replaced by machines and computers. Employees must be involved in this transformation at an early stage. This can be addressed through further training and a corresponding redesign of work organization.

- **Societal change:** Alongside the technological trends, political and societal change is also associated with this development. It is necessary to make the benefits of digital transformation available to society at large. This concerns data security, the emergence of new economic dependencies, the distribution of wealth, the design of the work of tomorrow, and also legal questions such as the liability of autonomously acting machines [9].

These points appear to pose a major threat to society as a whole as well as to the individual. In fact, however, these risks are offset by great advantages and positive developments: digital transformation offers people a great many opportunities to adapt to different activities – especially activities in which humans (still) surpass machines, such as empathy and creativity, i.e., the things that define humanity.

In a broader sense, technological change will go hand in hand with a new definition of work, and physical labor will be increasingly pushed back in favor of intellectual activity.

Examples include activities in care work, educational activities, and new creative professions, which will gain in importance. Through material prosperity, these activities will gain appreciation not only in ideal but also in financial terms. This is already a development of the past 50 years – as can be seen, for example, in the growing number of artists – and it will accelerate further over the next 30 years.

Technological change will therefore not only cost jobs but will also open up new opportunities for many people. In order for as many people as possible to be able to participate in the world of work, policymakers still need to set many courses – in the field of education as well as in tax matters, i.e., the redistribution of wealth [10].

2.6 Summary

The development of automation continues to progress steadily. The automation pyramid forms the basis for depicting the individual levels of action. The individual levels are connected to one another through various communication techniques and are the target of optimization in the Smart Factory. To master the future challenges, technological foundations and standards must be described and established.

Fundamental to future digital transformation is the shift away from a uniform, global control approach toward decentralized planning and organization in manufacturing. This is also reflected in the five paradigms, which represent the five core principles of Industry 4.0. This shift is necessary because the complexity in the corporate environment will continue to increase, and decentralized control can counter this complexity through local solutions.

The most important future challenge lies in translating the still comparatively abstract principles of human-centricity, resilience, and sustainability into measurable, operationally manageable concepts, while actively shaping the smooth transition from existing Industry 4.0 infrastructure to an extended Industry 5.0 practice.

Open questions remain regarding security, the liability of autonomously acting machines and the resulting consequences, as well as demographic change and the political changes that arise from advancing technologization.

CHAPTER 3

Technological Foundations and Standards

Learning Objectives

After working through this chapter, you will know ...

- ... which options exist for identifying objects.
- ... which description languages exist.
- ... what ontologies are and how they are used.
- ... what a digital object memory is.
- ... which types of human-machine communication exist.

Introduction

In the automation industry, the demand for optimized production processes and transparency is growing. For this reason, know-how in the field of automation and process technology is being combined with the development of standardized, intelligent software products. This enables optimized and transparent manufacturing processes that ensure high quality in production. The linking of hardware and software creates new synergies in the manufacturing process.

3.1 Identification of Physical Objects

To ensure real-time networking of physical objects such as goods and materials, these must be uniquely identified in logistical processes for tracking purposes. This is achieved through automatic identification procedures, so-called (Auto-)ID technologies.

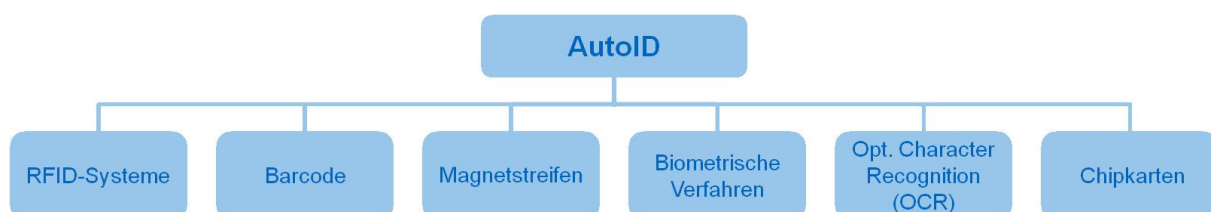


Figure 3.1: Automatic identification procedures (Auto-)ID technologies

Here, in particular **RFID technology** (Radio-Frequency Identification) and barcode technology play an important role in a smart factory.

”Radio-Frequency Identification is an automatic identification procedure. By using radio waves in the radio frequency range, it stores or reads information (usually an ID number) of a mobile data carrier by means of electromagnetic or electrostatic coupling between a read/write device and that data carrier” [11].

RFID systems consist of **transponders**¹ and networked reading devices. The transponder here is a radio tag that is attached to the object. Using the reading device, this object can then be wirelessly detected and read out. The reading device subsequently sends the captured data over the internet for further processing.

An alternative Auto-ID method is barcode technology. The barcode consists of a binary code composed of parallel bars and gaps of varying widths. Laser beams scan this code, and the pattern is recognized based on the different reflections of the light beams. Barcodes can be implemented as 1D codes (bar codes) or 2D codes, for example as a Data Matrix or so-called QR codes.



Figure 3.2: Barcode examples in 1D and 2D form

3.2 Formal Description Languages and Ontologies

Language in general is the set of elements and rules that serves mutual understanding within the corresponding systems. These systems mainly comprise alphabets, rules for the structure of expressions, sets of vocabulary, rules for the structure of phrases, etc.

”**Ontologies** are formal descriptions, expressed through a language, of relationships between concepts” [12]. In ontologies, the terms used to describe knowledge from a particular domain are defined. In addition to machine-interpretable term definitions, they also contain the fundamental concepts and relationships between the content.

If complex processes are to be made comprehensible to humans, structured and graphical elements (e.g. flowcharts) must be used for modeling. The challenge in describing a smart factory lies in implementing the necessary modeling and representation in a standardized way. This is defined, for example, by the **Object Management Group (OMG)**².

As an example, the **Business Process Model and Notation (BPMN)**³ was introduced, which enables the representation of complex business processes using simple pictograms and was designed for the factory life cycle (FLC).

Another independent association is the **World Wide Web Consortium (W3C)**. Among other things, it deals with the standardization and development of suitable description languages for so-called factory data reference models (e.g. for data management). In addition

-
- 1 A **transponder** is a radio communication device that receives incoming signals and automatically answers or forwards them. The term transponder is a compound of transmitter and responder.
 - 2 The **Object Management Group (OMG)** is a consortium founded in 1989 that develops standards for vendor-independent programming.
 - 3 **Business Process Model and Notation (BPMN)** is an industry standard of the Object Management Group (OMG) used for the graphical representation and modeling of business processes.

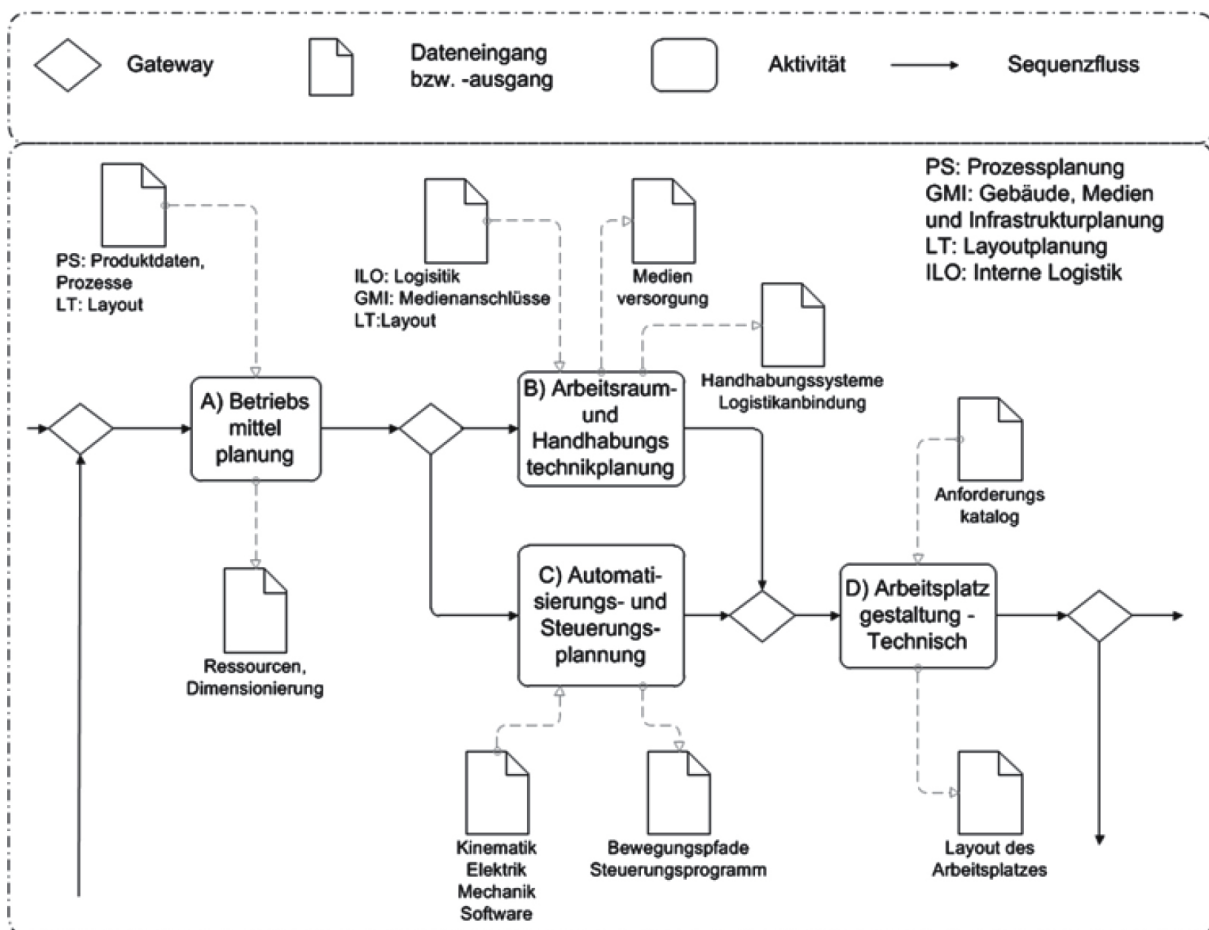


Figure 3.3: Model of a plant planning process with graphical elements according to BPMN

to machine-interpretable term definitions, fundamental concepts and relationships are also recorded.

Three examples of approaches for handling modeling are the **Unified Modelling Language (UML)**, **SysML (Systems Modeling Language)**, and also the **Web Ontology Language (OWL)**. UML was developed in the mid-1990s with the aim of unifying the large number of existing object-oriented methods. The goal is a holistic model whose central building blocks are structuring models and data management.

The complexity of a production company can be managed considerably better when a holistic model is used as a basis. Planning and continuous adaptation of all the structures and processes involved are made significantly easier through a holistic view. By dividing the factory life cycle into meaningful sections, it can be modularized and scaled, thereby providing data to the user in an application-appropriate manner.

Whereas UML and SysML are modeling languages based on a graphical programming environment, OWL is primarily intended as a classical programming language for semantic representation on the World Wide Web, consisting of various classes, properties, and instances.

3.3 Digital Object Memory

A **digital object memory (DOME)**¹ of a real object is a digital store that permanently records and retains all relevant information gathered throughout its lifetime.

Digital memories continue the idea of the Internet of Things in the field of information technology, namely equipping physical objects with an internet-based virtual identity. Mobile cyber-physical systems serve as the basis for such object memories, typically equipping real objects with sensor technology and processing capacity.

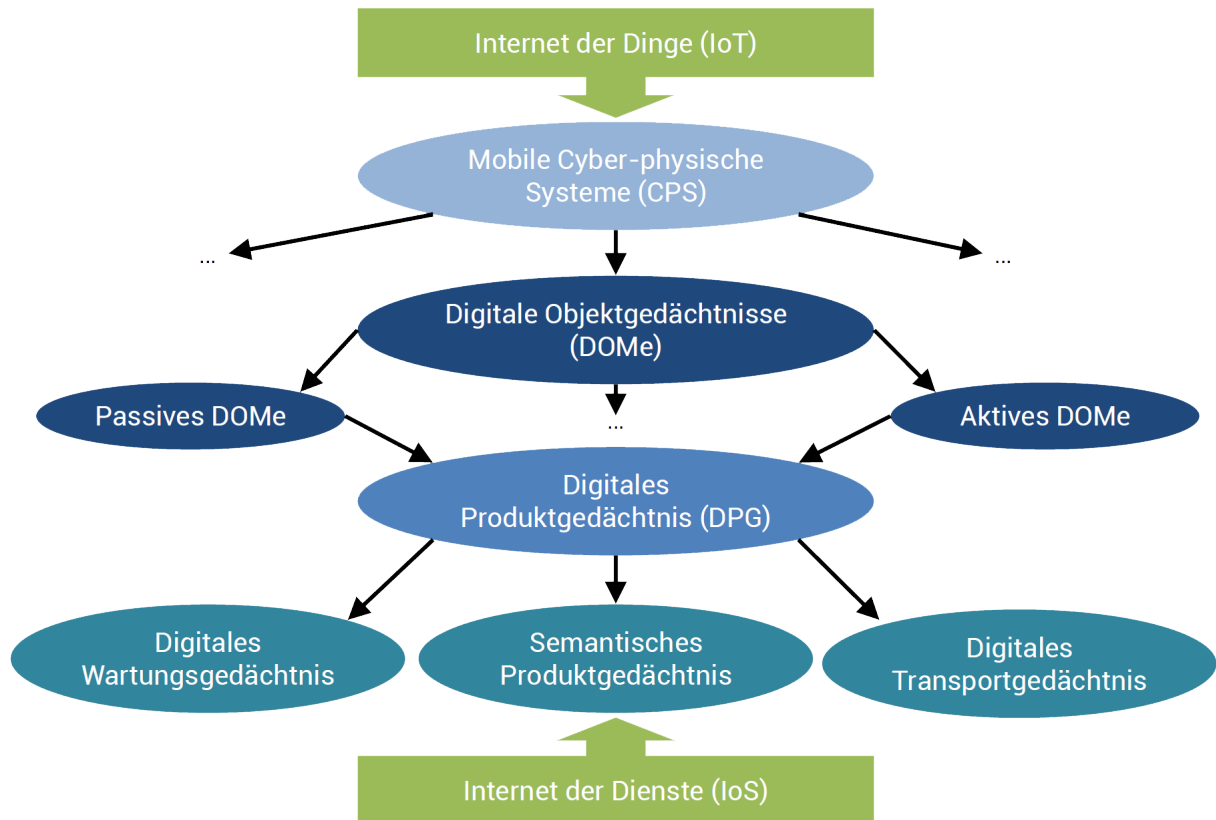


Figure 3.4: Object memories

Object memories can be distinguished, depending on the type of equipment, into passive or active memories.

- In the case of passive memories, identification is only possible at the physical object itself, through additional sensor technology and readout in the immediate vicinity.
- Active memories are those in which measurement and readout take place directly at the object.

In principle, these object memories can be used for any kind of object, for example for naturally occurring objects or even for user diaries.

Digital product memories (DPM) represent a subclass of object memories that was defined specifically for already manufactured products, whose task is to map a product's data throughout its typical product life cycle, from description and development through production to recycling.

¹ A **digital object memory (DOME)**, for Digital Object Memory is a data store in which all relevant information about a concrete physical object is continuously collected.

DPMs can be further subdivided into **digital maintenance memories**, whose task is to document maintenance intervals and periodic safety inspections. In the past, this was recorded on paper at the respective objects or production machines. Modern digital document management is gradually replacing this task.

Semantic product memories are a further specialization that contains not only machine-readable but also semantic, i.e. machine-understandable, information. Examples include product markings such as certification marks or seals of approval.

Digital transport memories contain all the relevant data required for transporting the object. For example, the temporal and spatial stations of products along the production path through the factory are stored for later traceability in case of quality problems. This data is also needed for material procurement or synchronization with the production machines.

By storing the object data, this information can be used by any application that has access to the corresponding descriptions in the relevant ontologies. This information can then additionally be combined with functions from the Internet of Services, such as maintenance services.

3.4 Physical Situation Awareness

The term »situation« is to be understood as the totality of the circumstances at a given moment. By recognizing this situation, machines or robots are able to understand certain behavioral patterns and respond to them accordingly. For this purpose, it is necessary that situations be captured for analysis through sensing of the environment (capturing objects of the real world, **Physical Awareness**). The technologies commonly used for this today are sensor fusion, pattern recognition, and situation awareness with the aid of situation maps [13].

Sensor Fusion

Sensor fusion involves combining several different sensors in order, on the one hand, to achieve redundancy in their data and, on the other hand, to increase the quality of the measurement data. This makes it possible to detect faulty sensors or signals and either correct them or at least draw conclusions about the state of the system.

Particularly in safety-critical systems, the use of multiple independent sensors is necessary to avoid incorrect decisions resulting from faulty measurements. In addition, certain situations can only be recognized at all through the combination of different sensors. An example is sensors for autonomous driving, where cameras and radar sensors, for instance, can only enable a safe overtaking maneuver together, through their interplay.

Pattern Recognition

Pattern recognition involves algorithms and systems for recognizing patterns from existing data. To this end, incoming data is compared with already existing patterns, recognized, and then assigned to pattern classes. Pattern recognition is essential for the physical world because, similarly to humans, situations can only be recognized by processing unordered quantities of data (measured values) into meaningful information, thereby forming the basis for a comprehensive situation assessment. Difficulties arise, for example, in **feature selection**, where experts must painstakingly configure the pattern classes using pattern recognition algorithms in order to correctly classify situations.

An as yet unsolved problem in pattern recognition is the shift of certain pattern classes over longer operating periods, known as »**drift**«.

Situation Awareness through Situation Maps

The information about physical reality provided by the sensors can be combined into a so-called »mental map« of the physical situation. This creates a physical and temporal representation of reality in which the recognized objects and environmental influences are stored as a dynamic occupancy map. In a smart factory, the location of all products should be known at all times. A challenge here is, for example, tracking object identity over time and changes in class assignment, which to this day is only possible through approximation.

3.5 (Partially) Autonomous Action and Cooperation

Cyber-physical systems must cooperate with one another in order to achieve given goals. The aim of physical situation awareness is a fully or partially autonomously operating production facility. Cooperating cyber-physical systems are able to negotiate courses of action from the recognized situations and implement them independently. Technologies must be capable of cooperative recognition, evaluation, and coordination of situations, or of negotiating courses of action. Among other things, situations must necessarily be recognized in real time using appropriate technologies.

These technologies mainly draw on the methods, algorithms, and approaches of **multi-criteria situation assessment** and **Artificial Intelligence (AI)**. These primarily cover the areas of decision-making, planning, and forecasting.

Since the goals are often unclear or contradictory, a particular challenge of artificial intelligence in the context of autonomous action is that it must be accompanied by a comprehensive impact assessment (e.g. societal impact).

The following machine-processable **basic technologies** are crucial for autonomous action:

Domain Model and Ontologies

The relevant knowledge required to enable autonomous action must be described in domain models. All information present in a **domain** is captured hierarchically by experts in standardized ontologies and can be exchanged across domains. Additional domain knowledge can be continuously incorporated into the models to extend the systems. Problems arise with vaguely formulated concepts, where decisions or goals can only be partially fulfilled.

Sensor and Actuator Technology

Sensors and actuators serve to observe and influence the physical environment of the CPS.

Sensors capture the physical or chemical properties of their environment, such as temperature, humidity, sound, or material composition, qualitatively or quantitatively, and convert the resulting measured values into digital form.

Actuators convert digital control values into mechanical motion or other physical quantities such as pressure or temperature, thereby influencing the environment. Challenges lie in ensuring the functionality, scalability, and self-configuring control concepts of sensors and actuators.

Communication Infrastructure and Platform

The growing prevalence of CPS necessitates a heterogeneous and hierarchically organized communication infrastructure. Communication networks can be distinguished into Personal Area Networks (PAN), Local Area Networks (LAN), and Wide Area Networks (WAN). These networks must be able to communicate with one another seamlessly.

In conjunction with mobile communication standards (3G, 4G, 5G, 6G), the current 5G standard in particular is designed to support real-time-critical services and spontaneous communication connections. To ensure reliable and seamless communication between the networks, the networks

are continuously monitored. The concept of Self-Organizing Networks (SON) could be one solution to these problems.

Efficient Parallel Processing Units

The ever-increasing integration of systems, and thus functions, on individual microchips offers increasingly fast and complex capabilities through parallelization. However, energy-efficient, space-saving, and cost-optimized solutions are increasingly required. Further challenges arise from the resulting complex security aspects (IT security) as well as ensuring correct functioning (e.g. in multi-core processor systems).

Distributed Stable Control Loops

Distributed control loops are control circuits built on independent, complex components that can be hierarchically and/or spatially distributed. The challenge is to control these loops stably and to avoid, detect, and correct connection failures. An example from the automotive sector is the decentrally controlled Electronic Stability Program (ESP), which forms a control network together with various sensors.

In addition to the technical questions regarding uniform communication interfaces and protocol standards, it must also be considered from what point a system acts autonomously or intelligently, and who bears the legal consequences arising from such action. This raises further scientific questions for other disciplines, such as law, which must be examined separately.

3.6 Human-Machine Interaction

Human-Machine Interaction (HMI) is concerned with how humans and automated systems interact and communicate with one another. This is by no means limited any longer to classical machines in industry, but also involves computers, digital systems, or devices for the Internet of Things (IoT). More and more devices are networked and carry out tasks automatically. The operation of all these machines, systems, and devices must be as intuitive as possible and must not overwhelm the user.

Cyber-physical systems support people in their actions and intentions, take over tasks, and thereby provide very high utility. At the same time, they partly decide and act autonomously, thereby also influencing human behavior and social processes. While these capabilities hold great potential benefit, mastering the acceptable design of human-machine interaction still remains largely an open challenge.

For communication between humans and machines to function smoothly, interfaces are required. This refers to the point or action at which a user comes into contact with the machine. Simple examples are light switches or the pedals and steering wheel in a car: when a switch is touched, the steering wheel turned, or a pedal pressed, an action is triggered. A system can also be controlled via text input using keys, a mouse, touchscreens, voice, or gestures.

Devices are controlled either directly, with users tapping on a smartphone's touchscreen or issuing a voice command, or the systems automatically detect a human need: traffic lights switch automatically when a vehicle crosses the induction loop on the road. With other technologies, the focus is less on control; instead, they supplement human sensory organs. An example of this is virtual reality headsets. In addition, there are so-called digital assistants: **chatbots**, for example, respond automatically to customer inquiries.

While the use of data from different sources leads to information gains, the following challenges in the human-machine interface must be overcome in order to provide suitable support to operating personnel:

- Generating useful information from the large volume of data for the various roles within a company,
- Presenting the newly gained information in order to make the processes transparent and comprehensible for people,
- Preparing the information for different display sizes (smartphone, tablet, monitor),
- Providing it across different operating systems (platform independence).

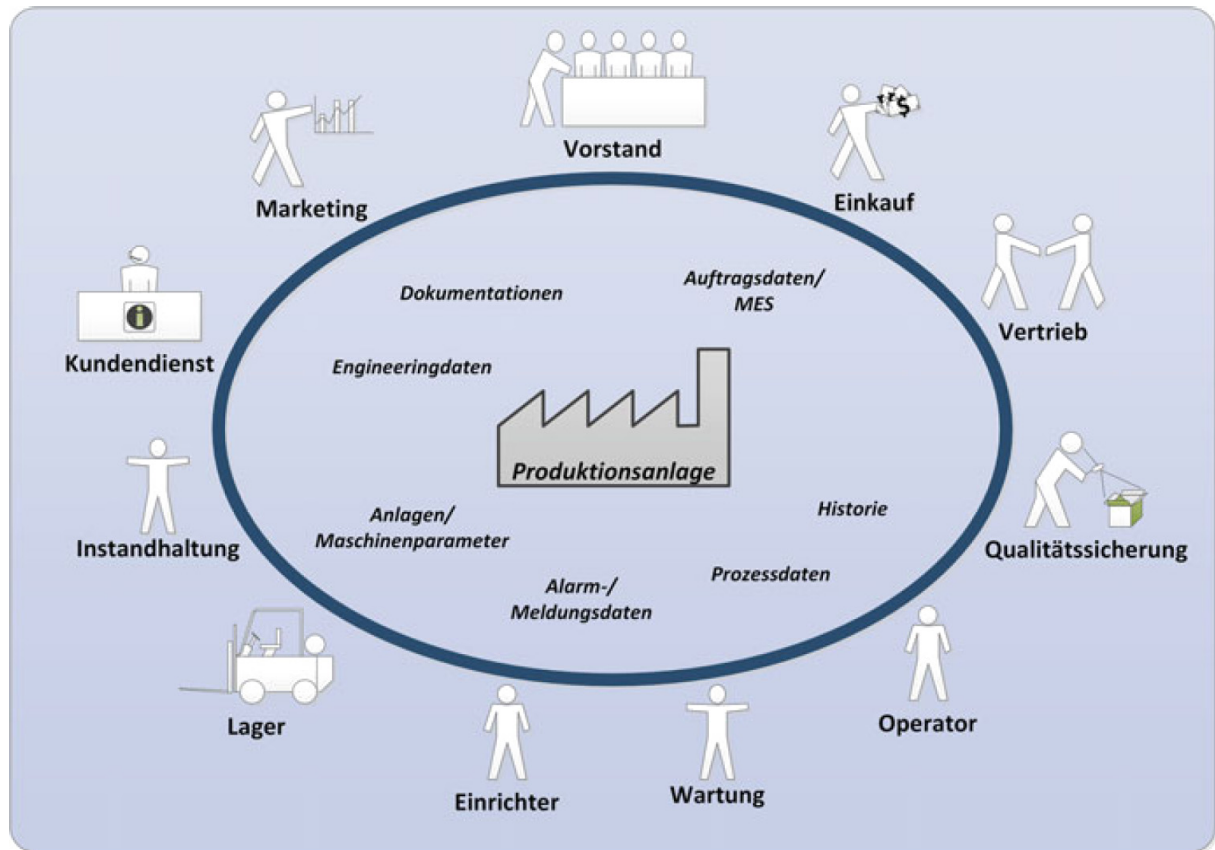


Figure 3.5: Roles and data classes in a production facility

Technologies

- Voice and gesture control
- Augmented Reality (AR)
- Social networks/information systems (chatbots)

Human-machine interaction, with voice and gesture control as well as virtual, augmented, and mixed reality, is by no means at the end of its possibilities. In the future, increasingly more data from various sensors will be combined (sensor fusion) in order to capture and control even complex processes as well.

At the same time, there will be increasingly fewer of the input devices common until now, such as remote controls, computer keyboards, or on/off switches. As computer systems, devices, and machines continuously learn and gain access to more data, they will also become increasingly similar to humans:

They will then be able to take over the function of sensory organs. A camera enables seeing, a microphone enables hearing, and clothing equipped with sensors enables a sense of touch.

3.7 Machine-to-Machine Communication

To enable the exchange of information between machines without human intervention, a machine-to-machine infrastructure (M2M) is required. M2M is closely linked to IoT, Industry 4.0, and the Internet of Everything. Areas of application extend across all industries and sectors, such as logistics, administration, production, security, and all types of remote access.

This is made possible by information and communication technology (ICT), although not all device platforms are currently standardized; rather, many M2M systems are designed for their specific requirements. Key components include a network of sensors, RFID tags, wired networks, and above all mobile wireless networks such as WLAN and 5G.

For the participating devices to be able to communicate with one another autonomously, it must be ensured not only that a shared vocabulary exists, but also that this vocabulary is semantically linked. Only this linking of terms makes mutual understanding possible between the participating objects. In the past, manufacturers, in order to secure their position in the market against competitors, developed their own proprietary local solutions and communication protocols for their end devices. The disadvantage of this isolation is that the devices are very difficult, if not impossible, to network with one another.

Various standardization organizations are therefore developing cross-cutting platforms in order to enable uniform communication among the rapidly growing number of end devices.

An example of a current standard communication protocol is **OPC UA (Open Platform Communications Unified Architecture)**¹.

This protocol has been developed since 2006 by the OPC Foundation, together with the 600 leading automation manufacturers such as Siemens, Rockwell, and Mitsubishi Electric. Currently, version 13 (IEC 62541-13, OPC Unified Architecture - Part 13) has been published, with the aim of unifying communication from the field level up to the cloud. In addition, semantic linking and security aspects are already provided for at the field level.

The foundation of OPC UA consists of many sub-standards, such as time-optimized Ethernet (TSN, Time-Sensitive Networking), and since 2018 has also been referred to as OPC UA over TSN, with a focus on real-time transmission.

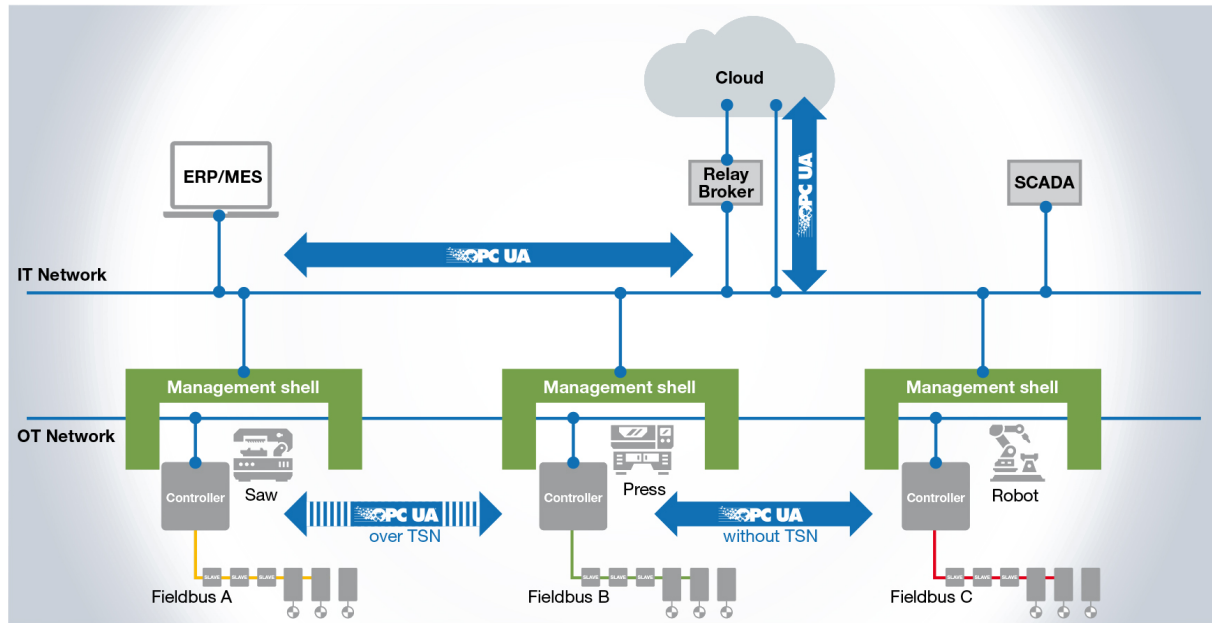


Figure 3.6: OPC UA, M2M standardization from the field level to the cloud

3.8 Summary

The technological foundations and standards on which a smart factory is based include the identification of objects, description languages, digital object memories, and situation awareness. Based on these technologies, partially autonomous action, and in the future also fully autonomous operation of the individual systems in production, becomes possible.

The consequences of autonomous action raise further questions, for example in the areas of society as well as of a legal nature. As processes become ever more complex, technologies for communication between humans and machines, as well as machine-to-machine interaction, are also becoming increasingly important.

Here it is of decisive importance not only that a shared vocabulary exists, but also that the terms are semantically linked, thereby enabling understanding. For example, the World Wide Web is currently being further developed in this direction.

¹ **OPC UA** stands for »Open Platform Communications Unified Architecture« and describes a collection of specifications that standardize communication in the field of industrial automation and M2M communication.

CHAPTER 4

Fundamental Concepts of a Smart Factory

Learning Objectives

After working through this chapter, you will know...

- ... what order-driven production looks like.
- ... how people in production can be supported.
- ... how products are becoming increasingly intelligent and how new services emerge.

4.1 Order-Driven Production

Market requirements continue to rise as a result of faster innovation and product cycles. Not only products, but also the means of production, are changing at an ever-increasing pace. Companies respond to these rising market requirements with **order-driven production (ODP)**. The order is placed at the center, and the goal is to assemble production resources optimally and dynamically.

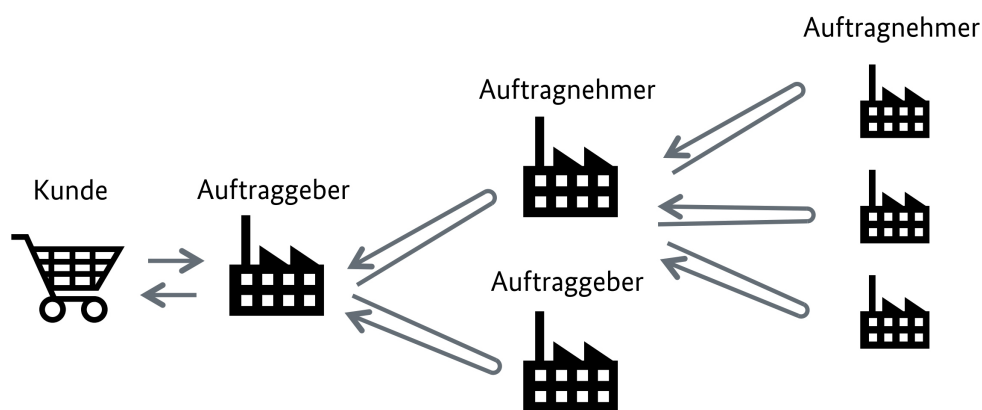


Figure 4.1: Order-driven production (ODP)

This is achieved by increasingly networking production (also beyond the boundaries of a single production site), enabling flexible responses to changing market and order conditions. In addition, this allows companies to improve response times and optimize manufacturing chains in terms of cost and time.

Order-driven production thereby relies on the standardization of all processes and the self-description of the means of production. The goal is automated order planning, allocation, and control in order to combine all necessary manufacturing steps and means of production with one another.

For example, companies can offer their own free manufacturing capacities to other production companies in order to avoid machine downtime and ensure utilization of their own machine fleet. Order fluctuations can thus be absorbed by both sides. The goal should be that production capacities can be integrated as easily as possible and that the production processes at both sites run autonomously in order to avoid delays at the interfaces. In particular, order allocation should be automated as far as possible.

Unexpected effects on value chains arise with respect to the competitive situation on the market. Due to the mutual interdependencies and roles in production (customer/supplier), mutual competitive situations can arise that must be taken into account. In addition, the markets of machine suppliers shrink, since the utilization of customers' machine fleets will increase.

These rather negative effects are offset by the flexible possibilities of reallocating production capacities. In addition, investment funds are not tied up, since external production modules can be drawn upon at short notice, and capacity utilization can be optimized.

4.2 Aggregation of Machine and Production Data

In the Smart Factory, machine and production data are stored and aggregated by virtual platforms. Intelligently combining this data yields insights for new services for customers or reveals optimization opportunities in production.

Data-Driven Services (Value Based Services, VBS)

In an increasingly networked industrial environment, new service offerings are becoming ever more important and are displacing well-established value creation processes. Examples can be found in the consumer sector, where streaming services are increasingly displacing the CD- and DVD-based media industry. Navigation devices, for instance, no longer offer only route planning but also additional services such as real-time traffic reports, which are made possible by the available user data.

At the center are IT platforms that, based on the use of products/machines, collect, analyze, and process the resulting data, thereby enabling new, individualized, and user-oriented services. These new services can, for example, be important for internal production:

- Maintenance appointments at the optimal time
- Timely provision of process parameters at the manufacturing workstation

The goal is to make consistent use of the data generated in the production process in order to further optimize production and to adapt quickly and flexibly—ideally autonomously—to changing influences (raw materials, manufacturing components). This not only optimizes production processes but, in the best case, also creates new elements of value creation.

In the industrial environment, two main actors operate who collaborate to a greater or lesser extent:

- Product provider (manufacturer of production plants / service provider)
- Customer (product user, operator of production plants)

With the introduction of »value based services«, a new actor emerges in the classical value chain, occupying the position between customer and product provider. This new actor, which can be either an independent company or the product provider itself, operates the IT platform that makes the resulting data available (platform provider). This provided data is passed on

as a raw material to the service provider, who can then offer new services to customers on that basis.

The service provider thus occupies the interface between the end customer and the product provider, which can lead to a sometimes considerable shift in the shares of value creation. From this perspective, it is essential for product providers to participate in this value creation process, to make use of appropriate platforms, and to offer their own services based on them.

The new actors create added value for all parties involved. Customers, for example, benefit from better maintenance and usage options through optimized adaptation of the product. Product providers gain better insight into customers' product usage and, as operators of the platform, can offer new services and increase customer loyalty, which leads to an increased share of value creation.

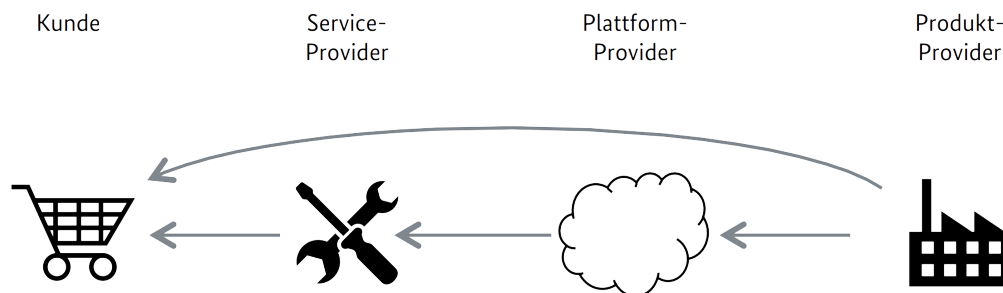


Figure 4.2: Data-driven services (value based services, VBS)

4.3 Supporting Humans in Production

Through digitalization, more and more people are able to cope with increasingly complex situations through information processing. The focus here is on human-technology interaction (HTI). Personal assistants in the form of smartphones or tablets and their apps help with this, for example, by automatically adjusting appointments to traffic conditions, pointing out the nearest gas station, suggesting restaurants, translating foreign languages, and much more.

In industrial production, too, digitalization means that people within the company are increasingly supported in complex and difficult situations, enabling them to reach assessments more safely and quickly. The secure handling of personal data always plays a decisive role here.

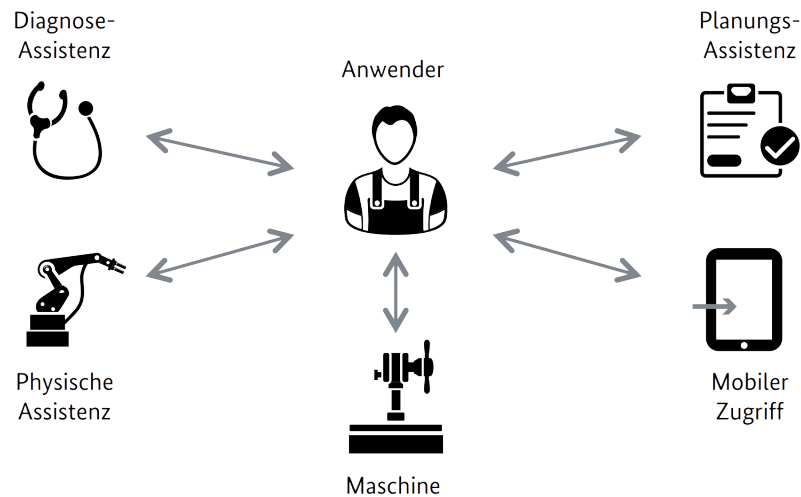


Figure 4.3: Human-technology interaction in production

Depending on the person's role within the company or in production, technology provides support in a wide variety of ways. For example, people are supported in heavy and complex physical work (capability enhancers). In addition, technology can contribute to analysis and evaluation for decision-making (assistance function). Support can, for example, be provided independent of location or through learning systems adapted to specific situations.

With increasingly flexible and intelligent systems, humans can increasingly focus on higher-level tasks. Physically demanding, monotonously repetitive tasks as well as highly complex tasks can be taken over by machines, thereby creating room for work content and higher-value activities such as quality management or safety, which are reserved for humans. This also strengthens employee motivation.

In addition, increasingly attractive workplaces for people in production are emerging, for example through individual ergonomic and linguistic adaptation. Motivation is increased, for example, through individual learning measures for new working methods, thereby avoiding frustration during further training. Additional support arises from the use of augmented reality in a wide variety of situations, from maintenance to quality management, or through remote access for experts in different locations.

The advancing technologization of production leads to changed work content across the entire value chain and thus to a new organization of work. A prerequisite for this transformation is that all employees in the company are informed transparently and comprehensively about the changes and made aware of the opportunities and risks involved. In particular, further qualification of employees with regard to IT-based decision support is becoming increasingly important for every individual employee and enables them to have more diverse and, at all hierarchical levels, greater freedom of decision-making.

As technologies increasingly conceal complexity and significantly influence people in their decision-making, the focus will increasingly shift to the consequences of these decisions. The discussion about changes in work content must therefore also be extended to include the distribution of responsibility.

In principle, digitalization opens up a flexibilization of human resources for the employer, since employees can be deployed without lengthy induction periods thanks to flatter structures. Through remote work, travel costs and travel time can be saved despite advancing globalization. Companies benefit directly from the attractiveness of the workplace and can thus more easily find well-trained employees.

4.4 Intelligent Products and Equipment

In order to operate a Smart Factory, not only CPS but also the products themselves and equipment such as machines are endowed with intelligence. An intelligent, or smart, object in general is an object that, through the integration of information technologies, goes beyond the object's actual purpose and thereby permits further functions and applications.

In the future, intelligent products will be manufactured directly from sensory materials. Their intermediate products likewise contain sensory properties and are interconnected via distributed networks. By capturing the diverse data generated during the manufacturing process, faults can be detected before they can have an impact on production (predictive maintenance). Data generated at the equipment can be transferred to the intelligent products in order to refine their properties.

The implemented intelligence acts as an intelligence amplifier and thus has a decisive influence on engineering, production, maintenance, and logistics. According to Wikipedia, these smart objects can be distinguished by scope of function into 5 areas:

- Identification and data storage
- Sensors for capturing the environment
- Actuators for influencing the environment
- Electronic data processing and decision support
- Communication and networking capability

In the industrial environment, a further three categories are distinguished. These require more or less embedded information technology. In detail, these are:

- Real-time data acquisition
- Decentralized data acquisition
- Networking

Above all, the networking of products and equipment enables data exchange between the objects involved. This makes it possible to plan and control production and its equipment in a timely manner and to optimize all corresponding processes. Communication between product and equipment leads, for example, to products telling the equipment—i.e., the machines—which programs a CNC machine should provide.

This is mainly exploited in logistics, for example to allow products to independently navigate through the factory to their production machines. Due to their embedded intelligence, intelligent products can naturally also continue to collect data later in the end-customer business and use it for smart services.



Figure 4.4: Intelligent products

4.5 Smart Services

Smart services are a direct consequence of consistent digitalization and thus the functional expansion of products through intelligent information and communication technology. The concept is very closely linked to the so-called Internet of Things (IoT). Products that were already equipped with embedded intelligence during their creation phase in the Smart Factory can of course continue to collect, process, send, and receive data after leaving production and being used by the customer. Because products can communicate via the IoT through platforms, entirely new kinds of services become possible that were previously inconceivable purely for cost reasons.

While in use at the customer's site, products permanently collect data and evaluate it. Enormous volumes of data arise, for example, from capturing environmental influences and user behavior. From this so-called Big Data, new knowledge emerges through algorithms, pattern recognition, and data fusion. Based on this „Smart Data“, the data of all customers can now be combined on a single platform. From this, knowledge-based services („Smart Services“) can be generated. Conclusions can be drawn from user behavior (real-time traffic jam reports) and offered to customers as additional services. Business models are thus expanded via IoT platforms, and expanded potential for value creation and employment can be unlocked.

In the future, a huge „universe of online services“ will emerge, in which actuators at the customer's site can directly influence the environment. Examples include apps that fully automatically control building technology such as heating, air conditioning, and lighting (e.g., lowering blinds). All areas of daily life will be permeated by these new intelligent services (business, leisure, culture, education, etc.).

The result will bring about a renaissance of online marketplaces [4]. This is due mainly to the newly developing semantic technology and cloud computing, i.e., IT platforms. The key to success is processing the enormous volumes of data—from information overload to precise, responsive behavior. The semantically described services can be reused, combined into value-added services, and flexibly expanded. Because this data and these services are available to all market participants on an open platform, it becomes very easy for them to develop new web-based applications, services, and business models.

The development of smart services enables companies and consumers to access this pool of services, content, and knowledge at any time and from anywhere. It also opens up enormous development potential for small companies such as startups and sole proprietorships, which can compete with large companies through innovative service modules and achieve rapid growth.

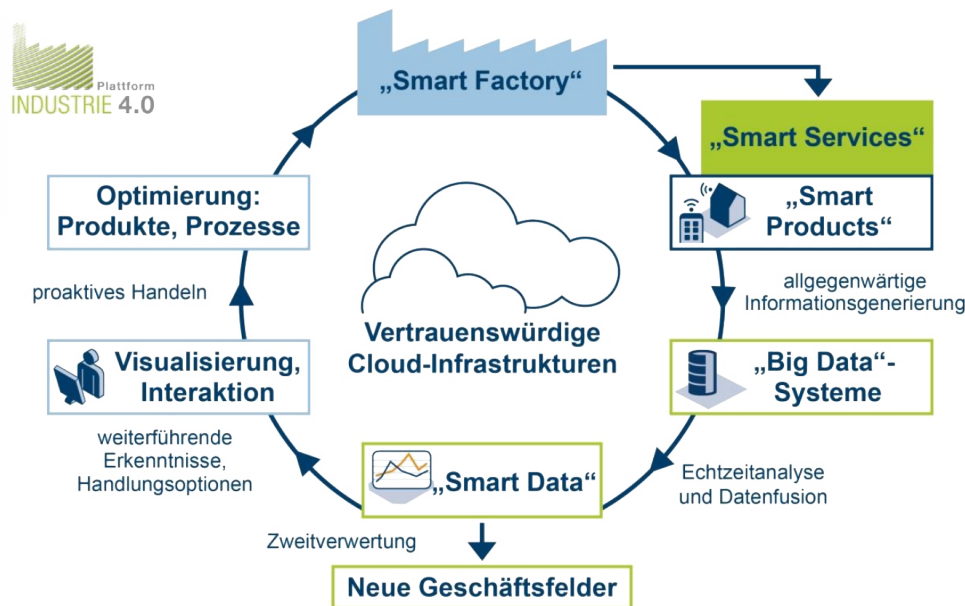


Figure 4.5: Smart services

4.6 Summary

In the future, companies will place orders at the center and thereby create autonomously acting production facilities that can respond flexibly and quickly to market and order fluctuations. Virtual platforms aggregate machine and production data. When this data is intelligently combined, it opens up insights for new services for the customer or reveals optimization opportunities in production for the entrepreneur.

For people in production, this means continuous further training and recognizing the opportunities and risks offered by the new technologies. Intelligent products and equipment have the autonomous factory as their goal and, through new services, will continue to push smart services onto the market. In particular, the new smart services, through easy access to data and data services via IT platforms, offer enormous growth potential, including for smaller companies.

CHAPTER 5

Reference Architectures

Learning Objectives

After completing this lesson, you will know ...

- ... what reference architectures are.
- ... what purpose reference architectures serve.
- ... which different standardization activities exist.
- ... how consortia from industry and academia contribute to standardization.
- ... what the RAMI4.0 reference architecture means

Introduction

Reference architectures are made available to customers (plant manufacturers and equipment manufacturers) in industry together with proprietary in-house solutions. This supports and simplifies, for example, the RAMI 4.0 integration of the reference architecture layers, communication, and the flow of information at the customer's site. With this implementation, equipment manufacturers can very easily forward data and information from the control level of field devices to higher levels.

Through so-called EDGE gateways, customers are thus able to achieve seamless networking of machines and plants all the way to the cloud. The advantages are scalability and, thanks to the open and platform-independent solution, an architecture that can be adapted to future requirements.

A further advantage for customers lies in the unchanged continued use of the customer's own systems at the fieldbus level, such as their own PLC controllers and the familiar, cost-effective IT network. This is possible because the diagnostic data is read out by the edge gateway from the field devices at the lowest level and prepared via browser or app for transmission to end devices such as tablets or smartphones.

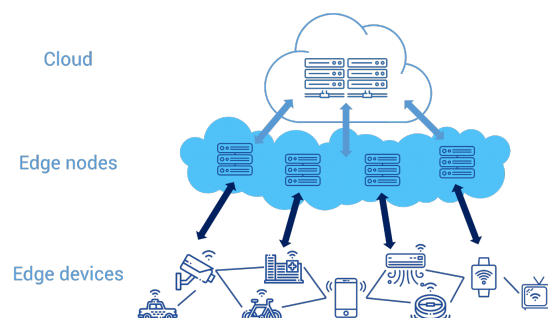


Figure 5.1: EDGE Gateways

5.1 Purpose and Characteristics of Reference Architectures

In computer science, the term **reference architecture** generally denotes an idealized model that serves as the basis for all further applications and special cases and represents the class of models and architectures to be modeled.

General reference architectures are required that unify the requirements of different manufacturers. »Reference architectures are kept independent of specific technologies and products in order to obtain the technically and economically optimal solution« [14].

However, committing too early to the architecture and functions of a single manufacturer severely limits the scope of action when designing the overall system.

Implementing the digital factory requires a high degree of end-to-end integration and standardization of methods, tools, and their interfaces. The basis for this is formed by the **reference models** and their subordinate reference architectures, which coherently describe the aspects of a system.

In doing so, the fundamental structures are determined and described, serving as the basis for a shared understanding within the professional community. It is not strictly necessary that identical solutions emerge for one and the same application, although reaching consensus should be the goal.

Harmonization and extension of the models and architectures can be achieved here through extensions and cross-domain measures. The primary goal, however, is to achieve a clear and unambiguous description of a matter across companies. Ideally, this even results in a worldwide standard for the relevant areas, although this cannot always be achieved. Different solutions for the same problems are the result.

To achieve this common basis, it is crucial to establish uniform communication (**interface management**) along the production process and to use uniform, redundancy-free data management. This can only succeed if all parties involved use the same communication structures and interface protocols, i.e., shared architectures.

To ensure all of this, there are wide-ranging efforts, for example by the federal government, professional associations (e.g., VDE/VDI), as well as other national and international cooperations among the companies themselves, to create a common basis for these communication structures.

Specifically, reference architectures affect the following areas of the production process for products, factories, and processes:

- Modeling
- Simulation
- Optimization
- Visualization
- Distribution and networking of data

5.2 Overview of Standardization Initiatives

The reference models and architectures for the smart factory are an essential component of the overarching Industry 4.0 initiative. There is a strong effort here to establish uniform structures and reference architectures in all areas of Industry 4.0, in order to defend and expand German and European excellence in production. This is only possible if all parties involved pull together, jointly establish rules for these architectures, and consistently pursue and implement them.

The German Standardization Roadmap »Industrie 4.0« sees the fundamental goal and basis for standardization initiatives in »harnessing the progress achieved, and expected in the near future, in information and communication technologies for production companies« (VDE: The German Standardization Roadmap »Industrie 4.0«. Version 5.0, as of January 2023).

Standardization and normalization are thus the drivers of innovation in the industrial environment and the Industry 4.0 concept, and require close cooperation between industry, research, and standardization bodies. Innovations are created in this way through

- Methodological grounding and functionality
- Stability and investment security
- Practical suitability and market relevance

The standardization process consists mainly of drafting a recognized document that contains specifications for general and recurring applications and is based on the agreement of all parties. The development of norms and standards takes place at national, European, and international levels.

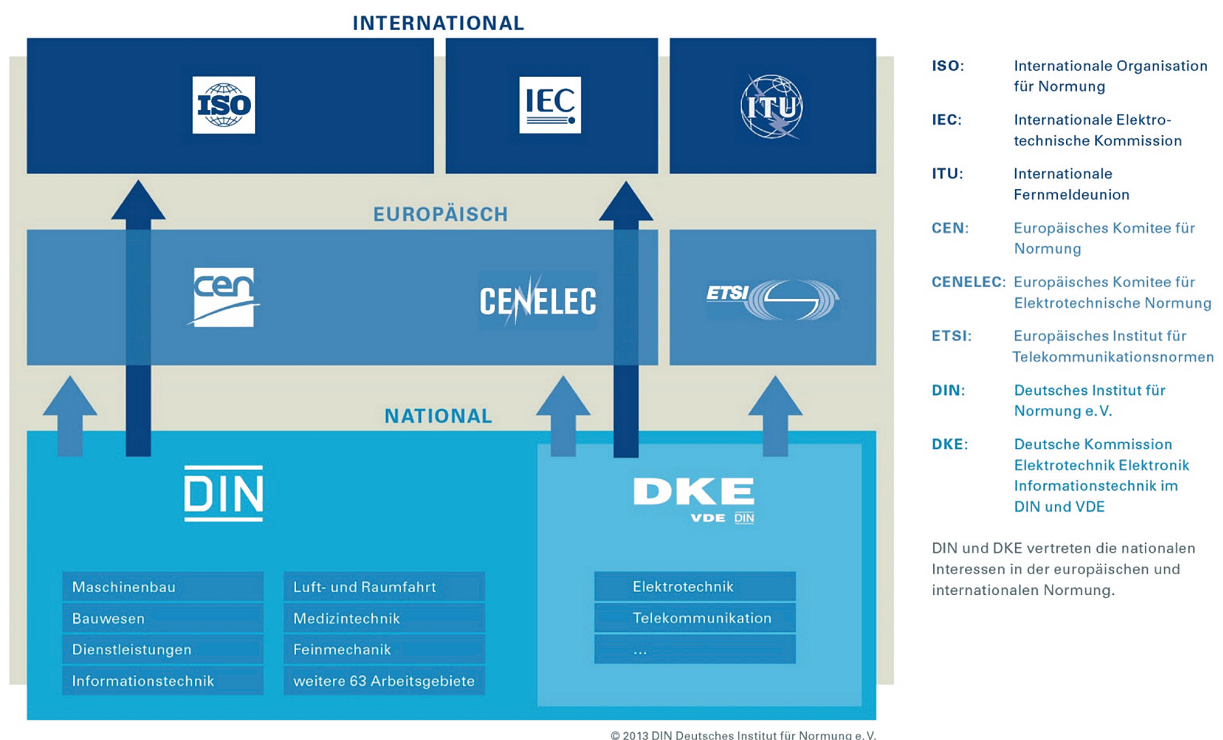


Figure 5.2: Standardization Initiatives

The standardization organization responsible for the Federal Republic of Germany is **DIN** (Deutsches Institut für Normung e. V., German Institute for Standardization). It represents German interests as a member of **CEN** (Comité Européen de Normalisation, European Committee for Standardization) and **ISO** (International Organization for Standardization) in European and international standardization.

The **DKE** (Deutsche Kommission Elektrotechnik Elektronik Informationstechnik im DIN und VDE, German Commission for Electrical, Electronic and Information Technologies) represents the interests of electrical engineering, electronics, and information technology in the field of international and regional electrotechnical standardization work and is supported by the **VDI** (Verein Deutscher Ingenieure, Association of German Engineers). The VDE currently has around 36,000 members who are active through individual and corporate members in specialist societies, state representations, and district associations. The society represents German interests both at **CENELEC** and at the **IEC**.

The **VDMA** (Verband Deutscher Maschinen- und Anlagenbau, German Mechanical and Plant Engineering Association) is Europe's largest industrial association and participates in the DIN standards committee for mechanical engineering (NAM). It is also the publisher of the VDMA specification sheets, the standardization procedure of the mechanical engineering industry.

Today, nearly 90% of standardization activities take place at the European and international level. DIN and DKE take the lead here in German participation and organize the respective national committees.

There are still further organizations and consortia worldwide that deal with future standards and proposals. Very often, these "**quasi-standards**" later flow into the final DIN SPEC and can be regarded as preliminary stages of standardization.

5.3 RAMI 4.0 (DIN SPEC 91345)

Within the framework of norms, standards, and concrete use cases, it is essential to develop common and uniform models for describing future requirements. This presupposes that a shared understanding of communication at the interfaces between the individual layers, together with their respective functionalities, is created. This is how the Reference Architecture Model 4.0 (**RAMI 4.0**¹) came about.

To remain holistic yet clear, RAMI 4.0 requires a representation that visualizes complex relationships in an understandable way. Using a so-called Smart Grid Architecture Model (**SGAM**), such relationships and details can be discussed relatively easily in a 3D representation. As a result, RAMI 4.0 has been adopted as a standard in the form of DIN SPEC 91345.

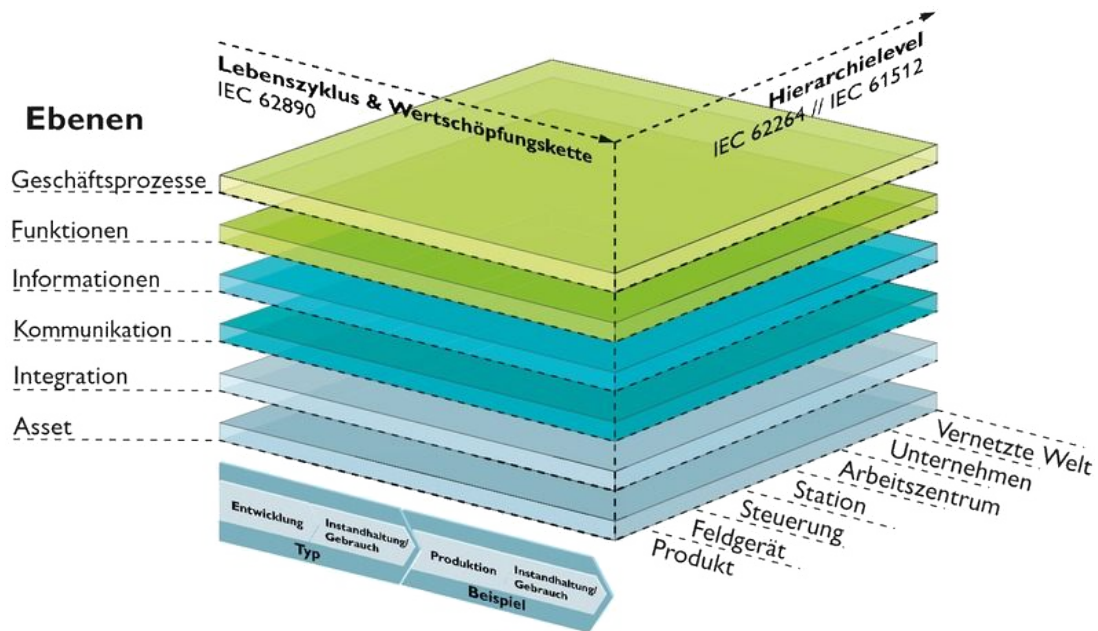
Based on three axes, the hierarchy levels of a networked production plant (vertical), the life cycle of plants and products (horizontal), as well as the IT representation (diagonal) of an Industry 4.0 component are described. Essentially, the hierarchy levels correspond to those of the automation pyramid.

Vertical axis (layers): Layers are used for the different perspectives, such as data image, functional description, communication behavior, hardware/assets, or business processes, of the production plant.

¹ **RAMI4.0** = is a three-dimensional map that describes how the topic of Industry 4.0 can be structured. RAMI 4.0 ensures that all Industry 4.0 participants can understand one another.

Horizontal axis (life cycle): Here, the product life cycle¹ is represented together with its value chains. This makes it possible to identify dependencies clearly, e.g., end-to-end data acquisition across the entire life cycle.

Diagonal axis (hierarchy level): Functionalities and responsibilities are distributed within the factories/plants. This concerns the functional hierarchy.



Quelle: Plattform Industrie 4.0

Figure 5.3: Smart Grid Architecture Model (SGAM), Reference Architecture Model RAMI 4.0 in 3D.

Some **objectives** of the RAMI 4.0 architecture are described in [15], Reference Architecture Model Industrie 4.0 (RAMI4.0):

- a clear, illustrative model as the reference
- use of existing norms and standards
- identifying and closing gaps in standards
- identifying overlaps
- minimizing the number of standards
- classifying relationships
- deriving rules

The basic building block is an I4.0 component, which consists of a real part (asset/machine) and a virtual/digital part (Administration Shell) and enables the interface to the upper architecture layers.

The RAMI 4.0 model offers many advantages. The architecture is service-oriented in design. All elements and IT-relevant components are combined into a layer and life-cycle model, so that complex processes can be broken down into manageable packages.

¹ **Product life cycle (PLC)** = a business administration concept that describes the process from the market launch or completion of a marketable product to its withdrawal from the market.

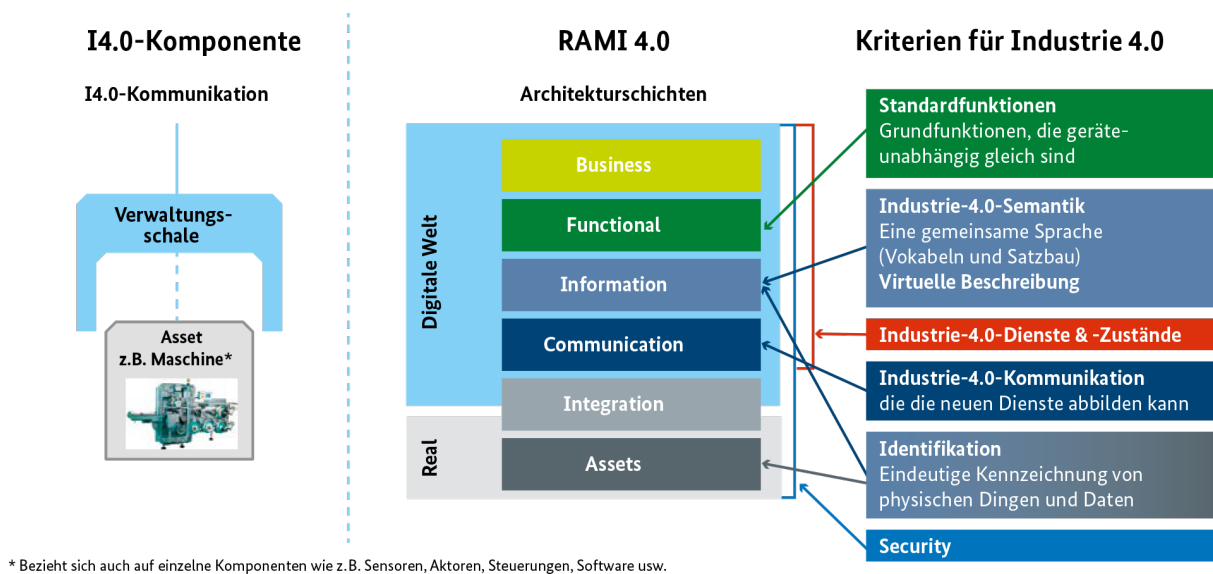


Figure 5.4: Reference Architecture Model Industrie 4.0 (RAMI4.0)

5.4 Summary

Reference architectures form the basis for standardization and normalization in the industrial environment and the Industry 4.0 concept. To achieve a shared national and international understanding of the networking, and thus the communication, of the CPS deployed, common models and architectures are needed.

To this end, standardization initiatives exist at all levels, which, given the wide variety of application scenarios and use cases, address these problems through norms and standardization and offer solutions.

The goal is always to develop a joint consensus in order to establish standards and norms, e.g., in communication and the IT landscape, across company boundaries as well. Greater flexibility and quality improvements are inconceivable without standards.

An example of this is the Industry 4.0 reference architecture RAMI4.0, which enables a very clear representation as a Smart Grid Model (SGM) and describes all facets and hierarchy levels.

CHAPTER 6

Smart Factory Engineering

Learning Objectives

After working through this lesson, you will know...

- ... how suitable methods and tools increase productivity.
- ... what User-Centered Design means.
- ... what is meant by Requirements & Virtual Engineering.
- ... what role the digital twin plays.
- ... how to combine classic and smart components.

Introduction

Today's production processes are characterized by increasing product variety and shortened product life cycles. In order to continue meeting high requirements in the future, e.g. with regard to response times («just-in-time production»), flexibility and reconfigurability, the approach of product-centric control holds great potential. Production resources are semantically described through information models, so that new variants at the production level can be realized solely by modeling the product to be manufactured during engineering. Parameterization at the equipment level is carried out entirely via the virtual twin. This significantly shortens the commissioning of new variants and processes!

The product determines its process! An in-house developed control system maps the requirements defined in the product model according to IEC 62264 onto the available production cells and thereby determines the logistical sequence. For this assignment, the heterogeneity and complexity of the (conventional) cell level is encapsulated by means of a developed OPC UA middleware - all functions are offered to the control system (MES) as a service. Locally, the process stations in the Smart Automation Lab access the assigned product model - the virtual twin - via AutoID (RFID). This integrates all information necessary for the production process, e.g. print image, assembly position or order assignment. In addition, cloud-based technologies, digital image processing methods for optical position correction, and approaches to human-robot cooperation are being pursued.

6.1 Classification of Different Engineering Tools

In the context of the Smart Factory, a tool is a computer-aided aid for the development of software, also called **CASE** (Computer Aided Software Engineering). Engineering tools relieve engineers of routine tasks and of the processing of data and information. The use of methods supports and simplifies changes.

However, tools are not miracle weapons, as tool manufacturers like to claim. Engineering tools cannot improve productivity by orders of magnitude, cannot replace independent thinking and careful work, and do not make quality management superfluous. The entirety of engineering tools (integrated CASE tools) can be classified into Upper CASE and Lower CASE tools:

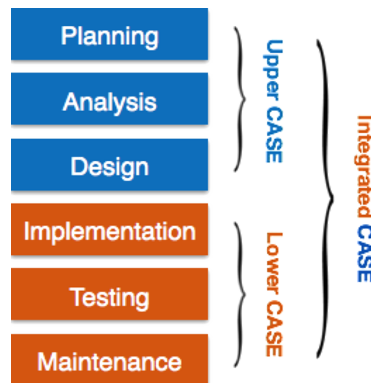


Figure 6.1: Classification of CASE tools

Examples of Upper CASE tools are:

- Specification and design systems
- Program design systems
- Program generators
- Analysis tools

Examples of Lower CASE tools are:

- Editors
- Compilers, browsers and programming environments
- Measurement and test tools
- Data management systems

The use of these tools can realize a productivity gain. However, when new tools are introduced, productivity initially decreases during the introduction phase. The reason lies in the necessary training, familiarization, and shifting of effort. The actual benefit, i.e. the break-even point, is only reached in the medium term. The introduction of tools represents an investment.

The major themes in engineering in the context of Industry 4.0 are:

- new technical methods (agile¹ development methods)
- legal issues
- new value chains and sustainability

To achieve sustainable efficiency gains in engineering and IT, suitable processes and coordinated tools must be used. The key to success is the combination of lean and simple processes with a systematic methodology.

Processes bring efficiency, transparency and discipline. Methodology creates systematics and support, and tools secure both sustainably. Tools help to specify requirements systematically and maintain them with discipline throughout the course of the project.

6.2 Virtual Engineering

The term engineering is defined as the creative application of scientific findings and mathematical methods for the development and production of technical artifacts (International Association of Engineers (IAENG) 2018). Engineering is often used synonymously with the term product development.

The Virtual Engineering approach aims primarily at linking the real and virtual worlds. Originally, support for the product development process was based on digital (three-dimensional) models. The focus can be on product development processes, industrial engineering, as well as other development processes for technical objects (such as buildings) and even for services. Virtual Engineering methods can be applied across industries.

The basis for these new engineering methods are technologies such as AR (Augmented Reality) and VR (Virtual Reality). Through the graphical representation of products in the early development phase, improvements can be achieved through iterative development cycles, without time- and cost-intensive prototype manufacturing.

In addition to classic development environments, AR and VR can also be used for employee training, for presenting prototypes to customers, or for the virtual commissioning of entire factories and plants. For example, it is possible to identify weaknesses in logistics and assembly processes through simulation before the production facilities are installed and set up on site.

¹ **Agile development methods** = approaches, primarily in software development, that increase transparency and flexibility and are intended to lead to a faster deployment of the developed systems. Agile software development is characterized by self-organizing teams as well as an iterative and incremental approach.

The goals of Virtual Engineering are:

- Enabling fast development cycles
- Obtaining early result feedback
- Strengthening early development phases
- Developing alternative product concepts
- Providing decision support when specifying the product

6.3 User-Centered Design

User-Centered Design (**UCD**) refers to a user-oriented design of interactive products. The goal of UCD is to increase the usability of products. For this to succeed, the end user must be placed at the center of product development. Accordingly, the customer's requirements and goals are at the center. Since around 1990, the term User-Centered Design has been used for this and has displaced the term Usability Engineering.

Many things in everyday life are not intuitive to use and leave users frustrated. In order to make product design user-friendly, manufacturers involve their customers very early in product development and gather experience throughout the entire development cycle.

To make matters more difficult, it is not only the end user who needs to be able to handle the product easily; it is often the case that people in, for example, higher-level positions also need to find their way around quickly.

Four basic recommendations for how a product should be designed:

- As simple a design as possible, so that intervention is possible at any moment
- Make things visible (concept of the system, alternative functions, effect on inputs)
- Simple design to recognize the current status of the system
- Pay attention to natural mappings, between intention and required inputs; between inputs and result; and between the visible information and the interpretation of the system

These guidelines place the end user at the center of the development process. Unfortunately, it is not enough to instruct development engineers that they should design products to be intuitively usable. Design guidelines are needed to guide the development process.

According to Norman [16], the following **seven concrete principles** are proposed for how to design a product in a user-friendly manner:

1. Use one's own knowledge as well as external sources to write user manuals that are easy to understand and, above all, are written before the design has been implemented.
2. Simplify the individual subtasks. The user's short-term and long-term memory should not be overtaxed. On average, a user can only remember 5 things at once. The user must always have control over the process.
3. Make things visible. The user should be able to recognize the purpose of an object by being able to easily identify the correct switches or buttons for performing an input.
4. The use of overview graphics significantly increases understanding.
5. Power of Constraints
6. Consideration of all possible user errors so that the user can undo the errors.

7. If nothing else helps, one should at least adhere to established standards.

UCD is particularly suitable for medium to large projects. They should have a certain degree of content-related and organizational complexity in order to be able to exploit the full range of possibilities. A project that is consistently carried out with UCD incurs lower costs than a project using other software development methods.

However, more effort is required in the early project phases, since the requirements and needs are gathered and described in much greater detail. In return, costs decrease toward the end of the project, and the probability of change requests is minimized, both during the ongoing project and during use after market launch.

6.4 Requirements Engineering

»Requirements Engineering is the systematic, disciplined and quantitatively measurable approach to specifying (i.e. capturing, describing and verifying) as well as managing requirements for a system« (Martin Glinz, Requirements Engineering, University of Zurich).

In simple terms, product development is centrally about understanding and describing what customers want or need.

»The accompanying process that tracks changes to requirements, manages requirements documents throughout their entire lifecycle, and plans, organizes and controls the core activities of Requirements Engineering is called Requirements Management«.

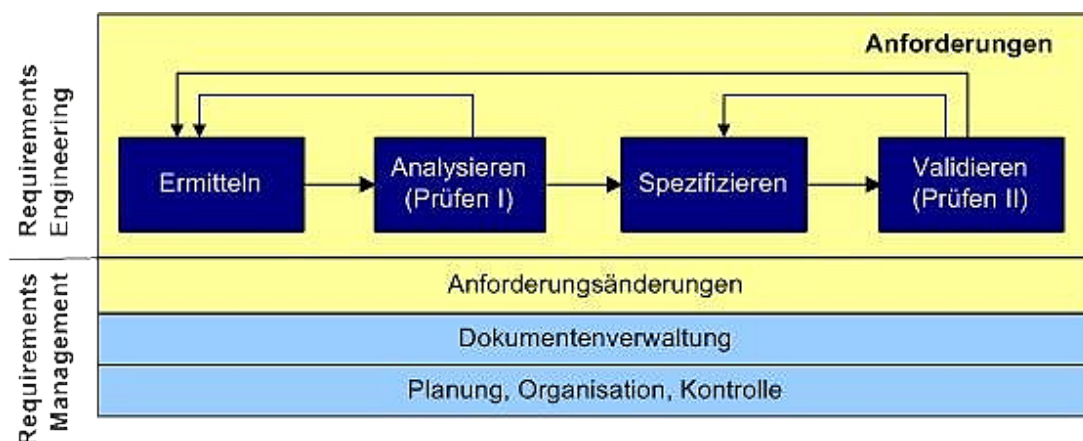


Figure 6.2: Requirements Engineering

In order to satisfy customers, it is necessary to adopt a human-centered view of the product before actual development begins. It happens often enough that a system is developed that is of no use to the customer or that the customer does not like. To minimize this risk, specifications and requirements are managed.

Of central importance is the effort to specify customer requirements. To do this, it is necessary at the beginning of a project to examine the actors in the environment of the planned system and their goals, actions and interactions in achieving these goals. From this information, the system context can then be derived, for example to delegate tasks, to provide the system with further information, or to obtain further information about the planned system. This is followed by the actual specification of the requirements.

The increased effort of systematically capturing these requirements is offset by the following advantages:

Costs are reduced:

- Lower manufacturing costs (by reducing costs due to errors)
- Fewer complaints and rework
- Lower maintenance costs Reduction of risks:
- Customer wishes are better fulfilled
- Forecasts for schedules and costs become more reliable

The characteristics of a good specification are documented in what is known as a **requirements specification document (Pflichtenheft)**, which should contain the requirements specification, an overview of the solution, and the elements of project execution, and should include the following

quality characteristics:

- **Adequate** - describes what the customer wants or needs
- **Complete** - describes everything the customer wants or needs
- **Consistent** - otherwise the specification cannot be implemented
- **Comprehensible** - for all parties involved, both customers and IT specialists
- **Unambiguous** - avoids errors caused by misinterpretation
- **Verifiable** - it must be possible to determine whether the implemented system meets the requirements
- **Risk-appropriate** - scope inversely proportional to the risk one is willing to take

A good specification process should be customer-oriented, carried out methodically and purposefully with suitable means, and should integrate the creation and verification of requirements into the production process.

6.5 Modeling

Modeling refers to the creation of a virtual image of a real or physical system. This is often also referred to as the **digital twin**. The digital twin creates a link between the physical and virtual worlds. This digital image of a real object contains all possible input and output channels as well as the entire specified functionality.

By changing the model parameters, the system can be very easily tested or virtually simulated for functionality and robustness before commissioning. In addition, feedback from the virtual model back to the real object can be achieved by collecting measurement data and forwarding this data. Actuators can then bring about changes in reality.

The virtual simulation of functional structures also enables targeted statements about the usage behavior of a smart product before the product has even been launched on the market.

Model Libraries

A major advantage of modeling arises from the implementation of what are known as module libraries. Here, very complex systems can be built hierarchically from the smallest function blocks or function models. These can be very easily inserted into new systems and reused. The result is a shortened development process.

Incomplete Functional Structures

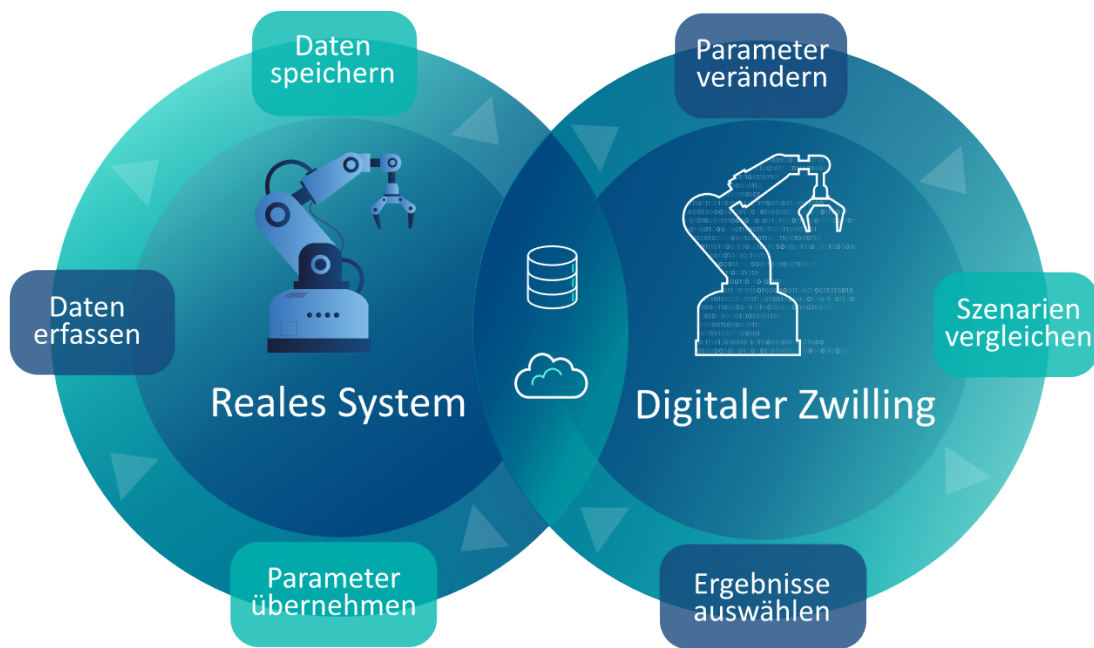


Figure 6.3: Digital Twin

There is a great need for action in choosing methods for defining incomplete functional structures. These are mainly descriptions of functions that can be dynamically modified both during the course of the development process and in the subsequent product lifecycle phases. In order to also meet dynamic customer requirements, this also includes incorporating customer processes into function development processes as well as their modeling.

6.6 Integration of Classic and Smart Components

The combination of classic and smart components and systems leads to intelligent objects, the »Cyber-Physical Systems (CPS)«. The classic components are supplemented with intelligent, internet-based services, known as **smart services**.

An example of a smart product is an intelligent vehicle that can be used by users via a car-sharing business model. In terms of their complexity, smart products can be individual products (smart car), networked product systems (several networked cars via a service platform), or very complex, cross-cutting product systems (entire mobility systems), known as systems of systems. Intelligent and networked products thus create completely newly arranged competitive conditions and even new markets.

The newly generated functions of these combined products can be divided into four areas that build on one another:

- Monitoring
- Control
- Optimization
- Automation

Depending on the application, the newly created intelligent products can cover all four of these areas. Each functional level can either serve independently as the basis for new functions or be used as a prerequisite for the next level of a functional area.

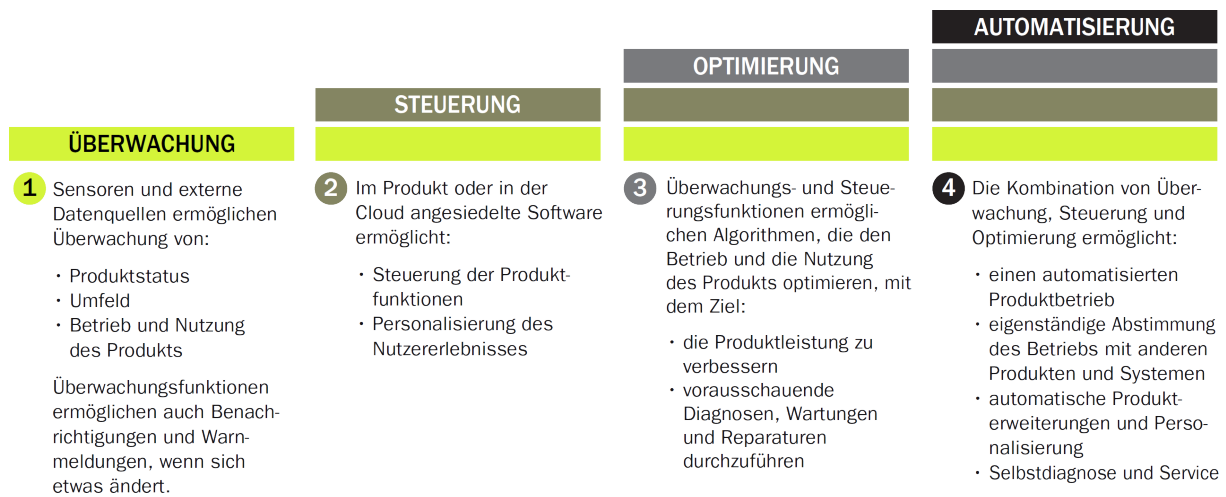


Figure 6.4: Possible uses of smart products

Networked products will not only change competition in the markets, but also the nature, work and structures of manufacturing companies.

The novel product solutions, which establish a previously unknown, close and direct as well as temporally unlimited relationship with the customer, are defined by three core elements:

- Physical components with mechanical or electrical parts
- Intelligent components, such as sensors, actuators, microprocessors, data storage, software/operating system, and a digital user interface
- Networking components, e.g. interfaces, antennas and networks, which ensure communication between the product and the cloud.

In many networked product solutions, the cloud will be a central element for the software/the product's "operating system", which hosts, for example, a big data database system, a rules/analysis engine, or intelligent product applications that monitor, control and optimize the product's functions.

6.7 Summary

Smart Factory Engineering is the starting point for the intelligent products of the future. Suitable methods and tools can enable faster development times and higher quality. However, this comes at the cost of increased initial expenditure, which nevertheless pays for itself over the course of the project. The advantages of Virtual Engineering, in which end-to-end digital product development is consistently pursued, are particularly evident in medium-sized and large projects.

Increasingly, more attention is also being paid to the eventual end user of the products. Here too, increased effort is required in identifying and documenting customer wishes. However, if requirements and specifications are systematically and thoroughly prepared, the susceptibility to errors is significantly reduced and customer benefit is significantly improved. All of this is further systematically focused through what is known as Requirements Engineering, which directs attention to the early development phases.

The modeling of digital systems represents the link to the physical world. By using module libraries, effectiveness can be increased, for example through the reuse of individual modules, leading to faster development times and more flexible development processes. The combination of classic products with smart, networked technologies creates completely newly arranged markets and even entirely new business fields, to which every company must adapt in order to remain competitive.

CHAPTER 7

Digital Twin

Learning Objectives

After working through this chapter, you will know ...

- ... what a “digital twin” is
- ... which key characteristics a digital twin has
- ... how the technology has evolved
- ... in which fields digital twins can be applied
- ... which opportunities and risks exist

Definition

A digital twin (German: „**digitaler Zwilling**“) is a digital or virtual copy of assets, processes, products, or services from our physical world. This so-called „**digital avatar**“ uses real-world data and reconstructs it at a virtual level. To create a digital twin, the physical object does not need to exist in its final version.

The goal of creating a digital twin is to gain new insights in order to improve processes, increase efficiency, or discover a problem before it occurs in the real world. The lessons learned from the digital twin can then be transferred to the original system with less risk and greater returns. All of this is made possible by the **Internet of Things (IoT)**, which collects data from the physical world and transmits it to machines for reconstruction.



Figure 7.1: Digital twin - Porsche Taycan.

7.1 History

The concept of “twins” in industry has existed since 2002, when **NASA** developed two identical spacecraft for the Apollo mission. The twin spacecraft was intended to prepare the astronauts before the mission, and during the mission, to facilitate the repair or maintenance of the spacecraft that was moving outside the physical monitoring range.

These “twin” systems enabled engineers and astronauts to work out solutions to problems during the Apollo 13 mission and to bring the spacecraft and crew safely back to Earth. Thanks in part to this development, NASA now uses digital twin systems to collect data, provide recommendations to crew members, and plan the requirements of future aircraft. The term digital twin was first taken up in 2010 [17].

Today, the use of digital twins has also spread into further fields of application and has become established as a new technology.

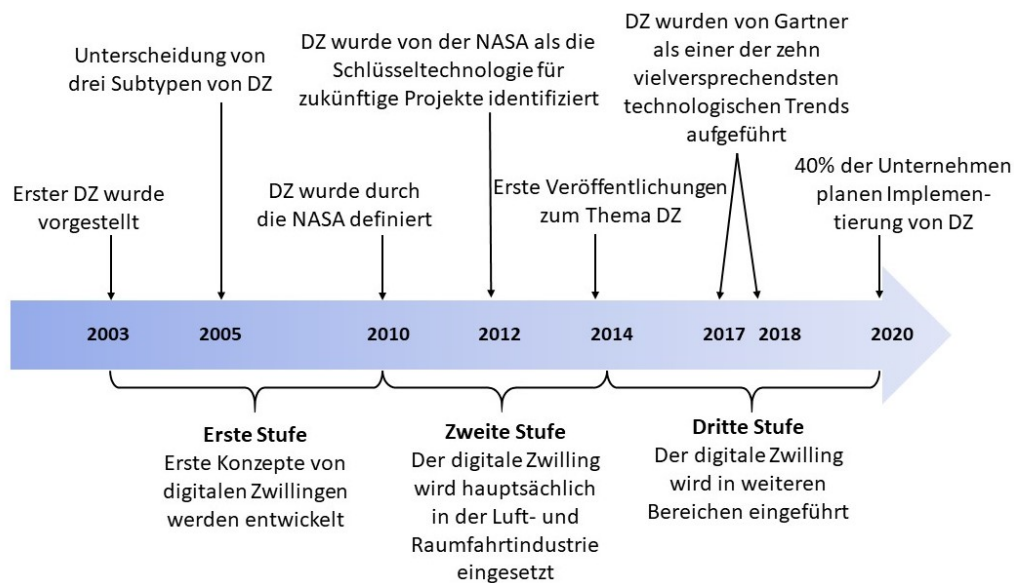


Figure 7.2: History of the digital twin [18].

As can be seen in Figure 9.2, the course of the development of digital twins can be divided into three stages.

Stage 1:

In the first stage, the concept of the digital twin was proposed by M. Grieves in 2003. Grieves, who is also a pioneer of **Product Lifecycle Management (PLM)**, defined this concept in three dimensions:

- physical entity
- digital counterpart of the physical entity
- connection linking the two parts together

In 2005, Grieves introduced a further idea, namely that the digital twin could be classified into three subtypes:

- digital twin prototype
- digital twin instance
- digital twin aggregate

Over the following five years, the technology continued to develop, laying the foundation for today's digital twins.

Stage 2:

In 2010, NASA published a definition and terminology for digital twins in the aerospace sector. In 2012, it was then announced that digital twins would become the new key technology for future spacecraft. Studies and research work then increased steadily, until in 2014 the first „**whitepaper**“¹ was published. With this step, “twin” technology was also introduced into further fields of application. [18].

Stage 3:

In recent years, many established organizations have attached great importance to the digital twin. Both in 2017 and in 2018, Gartner regarded the digital twin as one of the 10 most promising technology trends of the coming decade. Based on the current trend, it is therefore to be expected that the digital twin will undergo rapid development in the coming years. According to the **Digitalization Index**², as many as 42% of companies have now anchored the transformation of their business toward the introduction of digital twins in their strategy, or are already working with this technology.

7.2 Function

Digital twins can nowadays be used in many different industries and fields, such as the automotive industry, construction, or healthcare. In doing so, the complete product life cycle can be mapped, from development through production to aftersales. Objects connected to the Internet of Things are digitally and virtually replicated from the real world and enable simulation, testing, modeling, and monitoring based on the data collected by IoT sensors.

This data, collected in **real time**, is stored either locally in a decentralized manner or centrally in the cloud. The acquisition of data is the primary driver and the most valuable task of digital twins. Ultimately, this optimizes the use of real assets and products, and provides the ability to interact with them and evaluate „what-if“ scenarios.

In practice, two main concepts have become established for the use of digital twins [17]:

1. Using a digital twin to support product development in order to make design decisions, improve properties, or reduce the probability of damage in advance.
2. Using the digital twin for optimization after implementation, in order to improve maintenance intervals, components, or processes through the simulation of data.

¹ A **whitepaper** is an overview or collection of recommendations on a particular approach, standards, case studies, or research findings.

² The **Digitalization Index** is a survey of the level of digitalization of small and medium-sized enterprises. It provides information on how far digital transformation has progressed in various industries.

Within these different concepts, a further distinction is made between different types:

Digital product twin

The digital product twin is used in product development to commission a product virtually before manufacturing begins. The data collected during this simulated process helps to improve every aspect of the product, such as 3D/CAD models, test properties, or product characteristics. This is the most widespread type. It is usually used right at the beginning of data acquisition and delivers results that form the basis for subsequent applications.

Digital production twin

The digital production twin is used in the planning of factories, machines, plants, tools, and test programs. This type of digital twin supports, monitors, and optimizes production processes.

Digital performance twin

The digital performance twin is the digitized memory of a product. It contains production figures and information on throughput times, compliance with specified delivery times, and product quality.

Digital twin of the organization

The digital twin of the organization enables optimized, comprehensive planning of all processes within a company and increases data security and data protection [19].



Figure 7.3: Types of digital twin and their connection to the real product [20].

7.3 Structure and Optimization

Models that transfer products and processes from the **real** world to a **virtual** level have long been standard practice among today's industrial companies. However, they are usually only used prior to the actual use of the product and tend to be forgotten after implementation.

If, on the other hand, one looks at the concept of a digital twin, it goes beyond implementation. The digital twin accompanies the product throughout its entire life cycle. It also consists of three components: the real product itself, a digital image, and the networking of both areas. For the interplay and optimization of the real product and its digital image, the steps described in Fig. 7.4 are carried out one after another.

1. The starting point is basically the data already available, which can be collected, for example, through the use of the product.

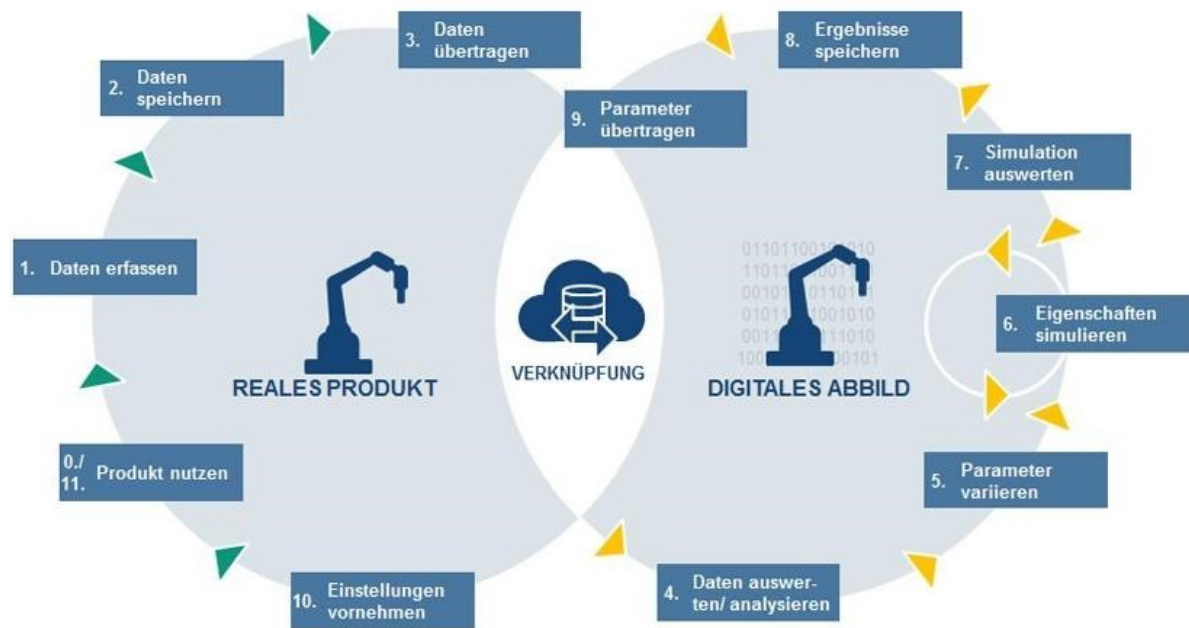


Figure 7.4: Establishing digital twins [21].

2. This data on the usage behavior of the product, machine, line, or entire production is stored in real time in a data space.
3. If this is a local storage location, transfer to a central location, such as a cloud, becomes necessary.
4. The collected data is then analyzed and linked together via variables in order to adapt it for the digital model.
5. The data collected from the real world is arranged in a parameter space, which provides room for maneuver for optimization through degrees of freedom.
6. The data is then modified by adjusting selected parameters so that, for example, new product properties result.
7. The product properties are in turn evaluated, assessed, and a recommendation is made for the real product. This can be repeated until the desired optimization is achieved.
8. The newly generated product properties are stored in the digital data space, just like the real data.
9. These „new,, data are ultimately transferred to the real world through the link between the digital image and the real product.
10. The data is adjusted so that the desired changes are achieved.
11. The product can now be used with optimized function.

The entire cycle is carried out continuously and thereby enables ongoing improvement of properties and an increase in the performance to be delivered. Many steps can be automated in the process.

7.4 Key Characteristics

A digital twin is characterized by the following five features:

Identity

A digital twin represents an individual physical asset. The digital twin can be simulated during development and exists until its real counterpart is decommissioned. However, thanks to the stored data, the information remains available even after the real image no longer exists.

Representation

The digital twin captures the essential characteristics of the real counterpart in a virtual format such as CAD or MES. The more precisely the digital twin is meant to resemble the real image, the more data must be provided and linked from different sources.

Status and Alerting

The digital twin reflects the current state of the real object in real time. Examples of such properties include location, process status, or speed. This enables the digital twin to trigger warnings and alarms on behalf of the object in the event of problems.

Context

The description of the operating context around the real image can also be displayed. This includes, for example, information about the installation, the people involved, owners and business partners, as well as notifications regarding maintenance and repairs to the object.

Interaction

Communication between the digital twin and the real object takes place continuously via the link between the two worlds. Only if this interaction can be reliably provided is it possible to use the benefits of the digital twin [22].

7.5 Challenges and Risks

The use of a digital twin offers opportunities and possibilities as well as challenges and risks that must be overcome. Depending on the design, type, and application of the virtual object, these have varying degrees of impact.

When creating **digital avatars** and using them, certain basic requirements must be observed. It is important that all the information a digital twin is supposed to process is brought together in a unified IT landscape. To this end, product designers, production planners, and application engineers must first develop a **shared understanding** in order to then be able to work toward a goal with a unified concept. All systems required in the various phases of the product life cycle, such as SAP, business portals, Windchill, various CAD programs, and databases, must be consolidated into a single access point in order to process the data meaningfully.

In each individual phase of the product development process there are **different boundary conditions** and optimization concepts. Each of these conditions must be considered for the individual phase in order to subsequently bring all phases together. The danger here lies in not knowing how these effects interact with earlier or subsequent activities [23].

Another disadvantage is the **unnecessary increase in complexity** of certain processes. Not every product or project requires a digital twin. In some cases, the increased effort involved would even extend the time-to-market. It must be clarified in advance to what extent the level of detail is necessary for the project and the associated digital twin, in order to then plan the necessary resources for the design.

The design of digital twins must be well thought out. A **methodical approach is necessary**, in which a virtual copy of the real world is created step by step according to agile procedural patterns. It is also necessary to consider the holistic use of the digital twin throughout the entire product development process. For example, interfaces, standardizations, and the flow of information must be defined in advance.

In order to be able to use digital twins in companies at all, **data experts** are required. It is not yet possible to do without human intelligence entirely. These data scientists are needed to create a working environment and to filter out which data, sensors, and information are actually necessary for creating a digital twin. They are also needed to recognize the connection to other internal Industry 4.0 tasks within the company and to build a bridge to them.

Another challenge is the **coordination and security of the vast amount of data** required to operate a digital twin. Digital twins typically have multiple stakeholders, each of whom should only have access to certain data. This poses risks to protected information, data privacy, and regulatory compliance. Great care must be taken to ensure that the right information reaches the right hands.

7.6 Opportunities

The virtual model of a product enables the **simulation and validation of product properties**. This makes it possible to discover characteristics, special features, and the behavior of the product in certain situations that could previously only be identified through elaborate testing. For example, is the engine still fully functional at an operating altitude of over 4000m above sea level? Such questions can be tested and optimized in advance using a digital twin [24].

Assemblies, entire machines, and production lines can be modeled in a virtual environment in such a way that **sources of error and disruption** can be identified and **eliminated** even before operations begin. The digital twin has the potential to intervene with methods and tools at every stage and to make changes. This saves time and costs, enabling highly complex projects to be completed efficiently in the real world with minimal use of resources [24].

Continuous optimization is achieved through the ongoing exchange between the digital twin and data from real operations. If an improvement to the current situation is identified, a change can be made directly to the component. This can go so far that the product can be improved in terms of usage to the point where it is perfectly tailored to the customer.

The use of digital twins results in **faster and more efficient product development**. It is increasingly possible to dispense with building real prototypes. This enables an earlier time-to-market, which ultimately brings the advantage of **time and cost savings**.

Cost savings can also be applied to the **early detection of product overhauls and maintenance deadlines**. By collecting data from real-world application, the digital twin can, for example, calculate the average wear of a machine and offer an optimal window for an overhaul. This reduces unpredictable machine failures and increases the availability of the equipment.

By combining several assets and their virtual images, entire ecosystems can be formed that can communicate with one another and enable improvements across the complete life cycle of product and machine. All in all, the use of digital twins leads to a **sustainably increased competitiveness**.

7.7 Examples in Practice

Digital twins can be applied to many industries. To date, the technology has been used predominantly in manufacturing and production, but it also has an impact on a range of other industries, which are illustrated below through examples.

Healthcare example

Digital twins can revolutionize both hospital operations and patient care. Creating a networked system of digital twins that represents an entire hospital and all the machines it contains can help surgical teams create a safe environment with faster patient care. Companies have already advanced to the point where it is possible to create digital twins of humans and organs. This enables complete simulations and test runs to be carried out before operations are performed on the person. This reduces potential risks as well as the mortality rate and helps improve patient outcomes.

Engineers at General Electric are working on developing miniature sensors that could monitor a person's vital information, such as heart rate, blood pressure, and oxygen levels. All this medical data would then be collected and linked to a digital twin. From this, the twin develops a more or less exact model of the person's body. The medical information could then be analyzed by software, which in turn would alert doctors to anomalies. With the integration of artificial intelligence, the software could even begin to provide diagnostic recommendations. This could improve care in hospitals, but could also give doctors insight into the well-being of their patients when they are not in the hospital.

Automotive example

It is predicted that by 2050, 12.4 million vehicles with highway pilot capability will be on the road [25]. The emergence of autonomous vehicles means an increase in new, interconnected technologies. Automobile manufacturers can create a digital twin of every single autonomous vehicle sold. This also makes it possible to analyze how a car behaves in its physical environment. Tracking manufacturing and production through to scrapping can also be easily monitored.

A large part of the sensor infrastructure is already present in newly launched vehicles, which constantly monitor a car's critical systems. When car owners go to the workshop and a mechanic performs a diagnosis, it is this sensor data that tells them what needs to be repaired. The next step is to set up systems that automatically send this data back to the manufacturer, analyze it, and carry out predictive analysis, which can help make the driving experience smarter and safer for car owners.

Example of urban infrastructure

To improve transport and travel systems in New York City, the government decided to create a digital replica of the city. By building an entire city on a digital platform, engineers are now able to develop digital road systems underground and predict precisely how this new transportation system will connect with the normal streets of New York City. Ultimately, using data on the average deterioration rate of a road in one of the city's busiest streets, they can make reliable predictions about how often the new underground roads will need to be replaced in order to ensure the safety of the new road system. By monitoring the use of these transport routes, the optimal time for maintenance work could in turn be determined in order to carry out the improvement [27].

Example in aerospace

In the aerospace sector, avoiding any potential risk is the top priority. But how can the average service life of a jet engine be predicted, and how does one know how the risk behaves if aircraft engines are not replaced in accordance with regulations?

By creating a digital twin of the aircraft engine, pilots can monitor the condition of the engine to see whether there is a potential risk of a fatal engine failure in the near future.



Figure 7.5: Creating networked systems in the automotive sector [26].

The same concept applies to all other parts of an aircraft. The positive aspect here is that simulation makes it possible to predict upcoming outcomes for the future without them actually occurring. The rule is: the larger the amount of data collected, the more accurate the resulting predictions. The U.S. Air Force now uses digital twins to apply warning rules that detect unusual wear or unforeseen damage development in engines and issue recommendations regarding further use and troubleshooting.

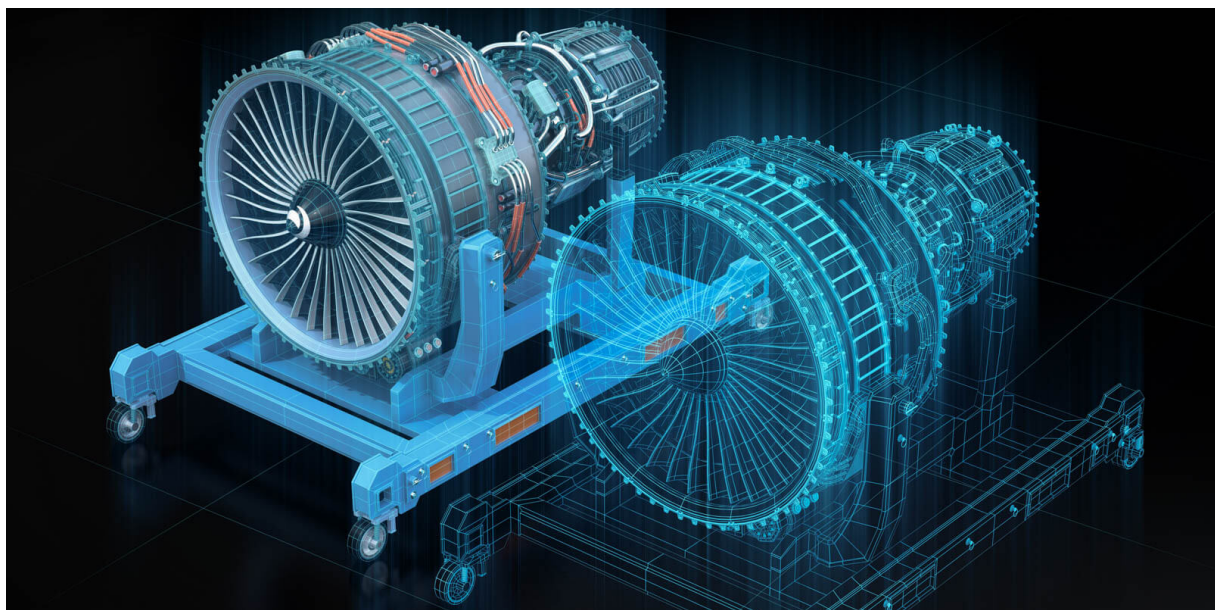


Figure 7.6: Digital twin of an aircraft engine [28].

7.8 The Future of the Digital Twin

At present, digital twin technology is still in its infancy. However, due to ongoing digitalization and increasing **connectivity**, this technology will spread into all areas of our lives. Above all, it is the immeasurable possibilities that will continue to drive this development forward. Because almost everything imaginable will be reproducible and digitizable in the future.

Now that digital twin technology has arrived on the market, developers can easily and accurately represent any physical object virtually by attaching special **control devices/sensors** that scan the object or product and carry out the model creation for them - this saves time and makes everything ten times more efficient. Digital twins will also improve insight into data and **decision-making** and ultimately help in the development of new business models.

For example, automotive engineers could use digital twins in combination with an analytics tool to analyze how a particular car is driven, in order to propose new features to **reduce accidents**. Yet other models could focus on potential new marketplaces for digital twins, interfaces, and correspondingly curated digital twin datasets [29].

Ultimately, this will enable stakeholders to improve the overall efficiency and costs of companies and will also change many aspects of work for employees. This potential will sooner or later transform and ultimately revolutionize the economy and industry as we know them today.

7.9 Summary

The first steps in the field of Industry 4.0 and the associated automation are being gradually introduced into our world today. This emerging digital world can be implemented through various approaches, one of which would be the introduction of digital twins.

Digital twins offer great opportunities in various fields, such as the product development process. The examples presented show initial approaches for integrating production knowledge or product twins in the respective fields. The continuous networking of digital twins is the key to further improving operational excellence and maintaining competitiveness within the industry.

The most important future challenge is to define a consistent framework for the holistic use of digital twins throughout the entire product development process. This should address the topics of a shared understanding of interfaces and standardization, as well as an efficient flow of information. In particular, this requires the integration of product designers and production planners in order to master new concepts in future research.

CHAPTER 8

Security Risks in the Smart Factory

Learning Objectives

After completing this lesson, you will know ...

- ... how to assess the security risks of a Smart Factory
- ... which recommended courses of action exist for a secure smart factory
- ... which guideline the VDMA proposes

Introduction

The rapid development of hardware and software in the field of industrial control systems (**ICS**) and Operational Technology (**OT**) has led to increasingly networked production environments in industry.

In the course of digital transformation and Industry 4.0, companies are using numerous basic digital-age technologies familiar from everyday life for improved and more efficient production: cloud computing, mobile apps, Big Data, machine learning, artificial intelligence, and many more. In doing so, however, they should always keep a critical eye on the cyber risks of these new technologies.

Security plays a particularly central role in companies that have a direct impact on health-related aspects, such as the food industry, the pharmaceutical industry, or medical technology.

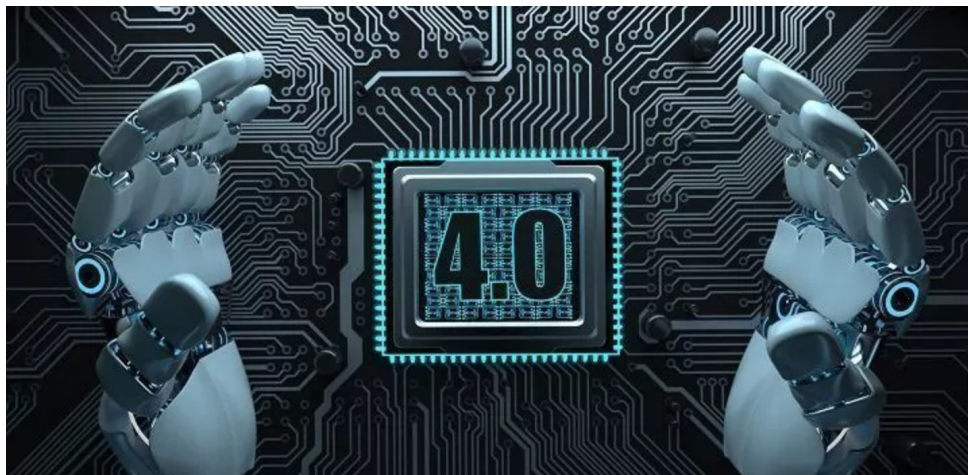


Figure 8.1: Security and Industry 4.0

The tendency toward the complete merging of devices on the internet leads to the fundamental dangers of attacks from outside or unauthorized use. Conceivable misuse scenarios range from unauthorized reading and spying on personal data, to the execution of undesired actions, all the way to sabotage and the destruction of entire industrial plants.

These new risks, created by networking, are to be reduced through appropriate security measures. The focus of these measures is mainly on technical as well as organizational IT security measures. In the Smart Factory, these risks can be divided into the aspects of **information security** (espionage, attack resilience) and **operational security** (sabotage, safety).

8.1 IT Security

One of the greatest technical risks within a Smart Factory is “IT security”. In recent years, cyber-attacks against companies as well as private individuals have steadily increased. Due to the insufficient development of data protection and data security, **cyber-attacks** cause substantial financial damage.

The spread of new business models and the use of cloud solutions are also being hampered by these emerging threats. CPS-based production systems create new security requirements because they exhibit a high degree of networking. They are connected to a large number of participating machines, IT systems, and automation components.

The technical system components partly act autonomously with time-critical data and information exchange, while at the same time multiple actors along the value chain are involved.

In manufacturing companies, a distinction can be made between the so-called **office IT environment** and the **production IT environment**. Whereas classic office applications in the IT domain are usually protected against hacker attacks and unauthorized access to the systems by currently maintained firewalls, production plants and machines are mainly controlled by older IT systems and are neglected in terms of IT security. The reason for production systems not being kept up to date with the latest IT security standards is the overriding requirement of availability. In other words, a production plant must not come to a standstill.

It is therefore common practice in companies not to modify a production plant once it has been installed (“**never touch a running system**”). Often, suppliers of production systems no longer offer security-related maintenance of the control systems after installation. In many companies, this means that, for example, companies no longer switch to the latest operating systems, and the plants become highly vulnerable to attacks over the network due to security gaps. The problems associated with systems that are still operational but outdated are also known under the term “**legacy**”.

In production plants, the use of access authorizations and password prompts is also often neglected in order not to make production dependent on individual employees.

To minimize risks in Smart Factories, it must be ensured that communication within the production plants is secured in such a way that no disruptions can occur and manipulation is ruled out. At the same time, the generated data must be protected against theft and misuse. Therefore, the individual systems and components must be protected against attacks such as denial-of-service attacks (**DoS attacks**¹) or malicious software.

The following security measures affecting all components of the IoT are necessary:

1 **Denial of Service (DoS)** refers, in information technology, to the unavailability of an internet service that should actually be available. Very often caused by overloading the data network (including deliberately).

- Powerful identity and access management
- Encryption of all data transmitted over the public internet
- Establishment of firewalls
- Robust systems for preventing attacks
- Software and patch management to remedy identified security deficiencies

8.2 Data Protection

In addition to the risks in the field of IT security, i.e., computer security, data security, and data storage (backups, etc.), the field of data protection is equally relevant. Here, the issue is not the protection of general data against damage, but the protection of personal data against misuse (»data protection is protection of the individual«). »The protection of personal data is based on the principle of **informational self-determination**. What must be protected here is privacy, i.e., personal data and anonymity must be preserved.

Every company is obliged to comply with the laws that are laid down exclusively for the handling of personal data in Section 1 of the German Federal Data Protection Act (BDSG).

In Europe, the European Data Protection Convention applies and was already agreed in 1981. This is a treaty under international law between the participating states that regulates the protection and cross-border exchange of personal data. Its official title is »Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention No. 108)«.

There is, of course, always the risk that employees who themselves have legal access to important data will misuse it. Preventing this is extremely difficult and cannot really be prevented by technical means. A healthy working atmosphere and respectful treatment of employees should be the top priority here, and not only to avoid misuse by dissatisfied employees.

8.3 Organizational Risks

In general, the field of the Smart Factory is almost exclusively reduced to the introduction of new technologies. Nevertheless, organizational risks must be weighed. If this is not done, there is a risk that the changes brought about by Industry 4.0 will lead to considerable problems, diminishing potential, and delays in the implementation of Industry 4.0.

The organizational structure of a company plays a particularly significant role at its highest hierarchical level. Here, management must define a clear strategy and a plan for digitalization in order to demonstrate an understanding of IT and processes. Organizational models must also enable cooperation as well as communication between all participants (people and machines) without hierarchical hurdles. In order to be able to prepare employees appropriately for the new tasks in a Smart Factory, management must set up a change team or a change department – in reality, this is often not implemented. This results in a kind of neophobia, whereby the new areas of responsibility are initially viewed very critically. In addition, declining employee satisfaction is frequently observed. For this reason, the change team must ensure that competencies as well as the qualification of the company's employees are in place.

The new technologies of Industry 4.0 change both the workflows and requirements of training positions as well as the job profile of individual activities as such. In production, far more skilled personnel are needed, while at the same time the proportion of employees with lower qualifications is decreasing, which can be attributed to digitalization. Companies currently face

the problem that many of their employees are not able to apply the new technologies because they do not understand the new processes. “Lifelong learning”, changes in work organization, team composition, and knowledge management will continue to play a significant role in the future.

8.4 Economic Risks

In addition to the technical and organizational risks, economic risks must also be weighed. The first challenge is the **lack of willingness to invest** – even though investments in Industry 4.0 are currently relatively inexpensive.

The biggest problem, however, is the **lack of transparency regarding the economic added value**. For example, companies in the mechanical and plant engineering industry must already consider today, at the time of investment, the requirements that products will have in the future. In principle, however, it can be stated that companies have long recognized and acknowledged the relevance of investments in Industry 4.0 across all sectors.

Companies must provide the necessary funds for investments and resources so that a Smart Factory can be implemented. Studies show that in Germany, as of 2025, **50-70 billion** euros are invested annually in Industry 4.0 applications ([Bitkom study Industrie-40.pdf, 2025](#)).

Another risk is the **loss of jobs** within a Smart Factory, since simple production tasks are being replaced by digitalization. However, the goal of introducing a Smart Factory should be to increase automation and reduce costs, not to reduce the number of jobs. The introduction of a Smart Factory requires the flexible deployment of personnel. There are already strong fluctuations in personnel requirements.

In the future, these fluctuations will be even greater. The requirements placed on production employees in terms of qualifications are also changing. Academic professions such as mathematicians, computer scientists, and professions in the fields of natural sciences and engineering, as well as non-academic professions in mechatronics and automation technology, must continue to be promoted in order to satisfy the high demand in the economy. Trainers must enhance the skills of trainees in the areas of Industry 4.0 technologies. Internal training and continuing education ensure that production employees remain up to date with the latest technology.

The introduction of the Smart Factory encourages employees to organize themselves and thus take on **more personal responsibility**. Consequently, the loss of certain business areas can be described as an economic risk. New market participants are displacing established manufacturers, such as bookshops and publishers, while startups open up new avenues for manufacturers.

In addition, the lack of support for research is a major obstacle. Germany invests well in research and funding; however, countries such as China and Japan, due to their greater willingness to invest, are performing better in the same area.

The identified economic risks indicate that, without taking these issues into account, a competitive and networked production environment cannot emerge.

8.5 Industrial IT Security According to IEC 62443 and ISO 27001

With the advancing vertical and horizontal networking of the Smart Factory (the convergence of IT and OT), classic office IT security is no longer sufficient.

While in office IT the confidentiality of data takes top priority, in production the priority is the availability and integrity of real-time processes. A production standstill or the manipulation of process data can cause devastating economic and physical damage.

While ISO 27001 forms the overarching, organizational management system for a company's entire IT and information security, IEC 62443 governs the concrete technical and operational security requirements for control systems in production (OT). Together they form a seamless security strategy in which ISO 27001 provides the strategic framework and IEC 62443 protects the shop floor from cyber-attacks.

IEC 62443 und ISO 27001

IEC 62443 und ISO 27001 sind komplementäre Standards für Informations- und Betriebssicherheit:

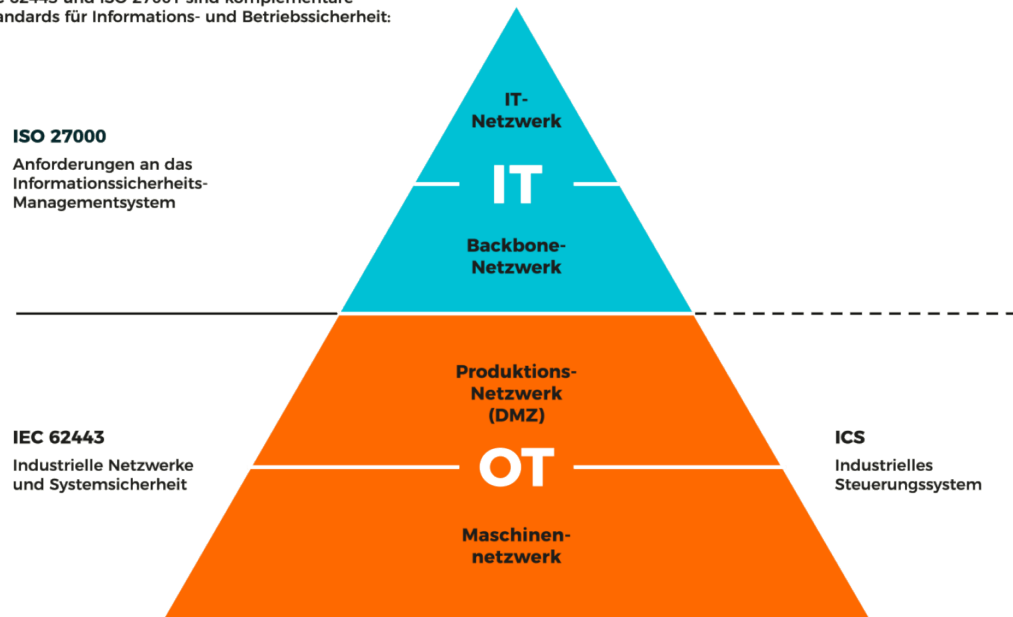


Figure 8.2: Industrial IT security according to IEC 62443 and ISO 27001

The Gold Standard: **The IEC 62443 Series of Standards**

The international IEC 62443 series of standards (“Industrial communication networks – IT security for networks and systems”) is the world’s leading standard for securing automation solutions (IACS - Industrial Automation and Control Systems). It addresses three key stakeholders of the Smart Factory equally:

1. Operators (Asset Owners): Must define security policies and processes.
2. System integrators: Must design and commission secure overall installations.
3. Device manufacturers (Component Product Suppliers): Must integrate “security by design” into hardware and software components (e.g., controllers).

The Architecture Concept: Defense in Depth (layered defense) Since there is no such thing as one hundred percent security through a single measure, IEC 62443 requires a multi-layered security approach. If one line of defense fails, the next takes over. The concept is organized into three main pillars:

1. Physical Security:
 - Access controls for server rooms, control cabinets, and production cells.
 - Deactivation of unused physical interfaces (e.g., USB ports on IPCs) to prevent malware from being introduced on site.
2. Network Security & Zone Model:
 - Segmentation of the factory into logical “zones” with the same security requirements (e.g., separation of the field level, control level, and cloud connectivity).
 - Definition of strictly controlled interfaces (“conduits”) between the zones. All data traffic (e.g., via OPC UA) must pass through firewalls and deep packet inspection (DPI).
3. System Integrity:
 - Hardening of the operating systems of machine controllers (disabling unused services).
 - Continuous patch management and the use of intrusion detection systems (IDS) that detect anomalies and unauthorized access in the OT network in real time.

Central Key Concept: Security Levels (SL) The standard defines four security levels against which components and systems can be certified:

- **SL 1:** Protection against unintentional or accidental violation of security.
- **SL 2:** Protection against intentional violation using simple means (hobby hackers).
- **SL 3:** Protection against intentional violation using sophisticated means (organized crime).
- **SL 4:** Protection against intentional violation using sophisticated means and specific industry knowledge (state-funded attackers/sabotage).

8.6 Recommendations for Action by the BMWi

The then BMWi, the Federal Ministry for Economic Affairs and Technology (today BMWK, the Federal Ministry for Economic Affairs and Climate Action), made it its premise to support German industry on questions of security and defense against national and international attacks.

Furthermore, there are facilities in the economy whose impairment could pose dangers to the life or health of the population, to public security or order, or to the defense readiness of the Bundeswehr. A particular danger can also always arise from persons who work at these facilities.

The BMWi has established the initiative »IT Security in Business«. The initiative is intended to raise companies' awareness of the topic of IT security and support them in improving their level of IT security. Together with IT security experts from business, academia, and public administration, the ten most important points were compiled.

1. IT systems, mobile devices, and services should always be kept up to date and regularly checked for viruses.
2. Security policies and IT system contingency plans should be created in an easily understandable manner.
3. Use of qualified electronic signatures, encryption for sensitive data, and care when sending and storing data.
4. Use of a clear and qualified authorization system.
5. Regular, spatially separated data backup with a view to availability at all times.
6. Cooperation only with trustworthy IT service providers or software vendors that operate according to European data protection standards.
7. Business-critical data in the cloud only in organizational forms (e.g., private cloud) that offer enhanced security, including encryption.
8. Always use encrypted transmission paths such as **VPN**¹ and proxy servers that do not store user data for IT devices.
9. Provide employees with guidelines on the use of social networks.
10. Regularly train and raise staff awareness of the topic of IT security.

Further information and free resources on the topic of IT security can be found on the website www.it-sicherheit-in-der-wirtschaft.de.

8.7 VDMA Guideline According to IEC 62443

The Verband Deutscher Maschinen- und Anlagenbauer e. V. (VDMA - German Mechanical and Plant Engineering Association), with its roughly 3,200 members, is the largest network organization and an important voice for mechanical engineering in Germany and Europe. It offers its members and companies support in areas including vocational training and education, research, innovation, law, standardization, and general questions of social and economic policy.

¹ **VPN** = Virtual Private Network, a virtual private communication network

From this association, the VDMA Forum Industrie 4.0 emerged, formed from an interdisciplinary team of VDMA experts who see themselves as partners and service providers. They support member companies in the fields of action relevant to Industry 4.0.

For companies pursuing Industry 4.0, IT security is vital to manufacturing processes. Above all, the autonomously running data exchange must be secure and reliable. In doing so, actors must be identified in a controlled manner, and all knowledge about products, machines, and plants must be protected.

Since classic IT security concepts fail in production, the VDMA has modernized its practice-oriented security guideline. The focus today is on the implementation of the international IEC 62443 series of standards.

The VDMA guideline breaks down the complex standard for mechanical engineering into practical steps:

1. **Defining the distribution of roles:**

The guideline governs the clear division of responsibility between the machine manufacturer (ensuring security by design), the system integrator (secure networking of the overall plant), and the factory operator (definition of security processes in operation).

2. **Risk assessment and zone model:**

Plants must be divided into logical security zones in accordance with IEC 62443. Critical controllers at the field level are strictly isolated from administrative networks or cloud connections. The connecting channels (conduits) between these zones are protected by firewalls with deep packet inspection (DPI).

3. **Embedding cybersecurity in the product lifecycle:**

Security functions must be actively maintained from development through commissioning and on through decades of machine maintenance. This also includes hardening systems by disabling unused services and physical interfaces (such as USB ports on machines).

8.8 Summary

In the field of security in the Smart Factory, many areas are affected by possible attacks and risks. Especially due to the connection to the public internet, espionage and sabotage from outside will repeatedly become a problem. In addition, internal employees are, of course, also repeatedly involved in criminal acts. The then BMWi (today BMWK) has made it its premise to support German companies in securing their IT and communication environments.

The result is concrete recommendations for action that companies should implement in order to minimize the risks of espionage and sabotage. The Verband Deutscher Maschinen- und Anlagenbauer offers a very detailed guideline for the rapid implementation of Industry 4.0 ideas.

Here, companies are given tools to drive forward, through interdisciplinary workshops within the company, the generation of ideas for innovative products as well as for improved production environments. Implementation can then be carried out immediately in projects on site.

Securing a Smart Factory requires a shift away from purely isolated IT protection measures toward a defense-in-depth approach. While ISO 27001 provides the organizational framework for the entire company, the modernized VDMA guideline, based on IEC 62443, provides the concrete, technical guidance for operational technology (OT) in modern mechanical and plant engineering. The BMWK's recommendations for action provide the overarching political foundation for this.

The implementation of Industry 4.0 does not succeed solely with the help of the federal government, but also requires the support of companies. It is necessary for companies and their management to be made aware of the topic of the Smart Factory and digitalization and to recognize the urgency of change in production. If this is not taken into account by German companies, this neglect will have a negative impact on Germany's competitiveness in the coming years.

CHAPTER 9

Sustainability in Production

Learning Objectives

After working through this chapter, you will know ...

- ... how the term sustainability is defined
- ... which different types of sustainability exist
- ... what the opportunities, benefits and goals of sustainable production are
- ... which concepts and methods exist for sustainable production
- ... why Siemens should be a pioneer for many companies

9.1 Sustainability

The origin of the term "sustainability" (German „Nachhaltigkeit"; English "**sustainability**") can be traced back to a way of thinking and acting in forestry. According to this idea, the future prospects of subsequent generations should not be worsened by the actions of the current generation.

Moreover, from a societal perspective, sustainability describes a policy whose goal is the equal and equitable **safeguarding** and **improvement** of ecological, economic and social performance. It is significantly influenced by the technologies that are available and deployed. In this respect, technology constitutes a decisive influencing factor.

By applying the principle of sustainability, negative events such as climate change, economic crises or social unrest are meant to be avoided or their extent reduced. Goals at the societal level are: [30]

- Preservation of the ecological basis of life and natural resources
- Securing and expanding prosperity,
- Preservation of social peace and health.

The Missed Opportunity for Sustainability

Looking at the figure below 9.1, it can be seen that as early as 2015, the equivalent of 1.6 Earths were mathematically needed to sustainably meet demand. That year, humanity had already used up all of its resources for the year by August 13. In 1987, by comparison, Earth Overshoot Day still fell on December 19. By 2050, demand is even expected to triple. Humans consume more

than can regrow or regenerate. At the same time, consumption keeps rising, the global middle class is growing, and emerging economies are catching up.

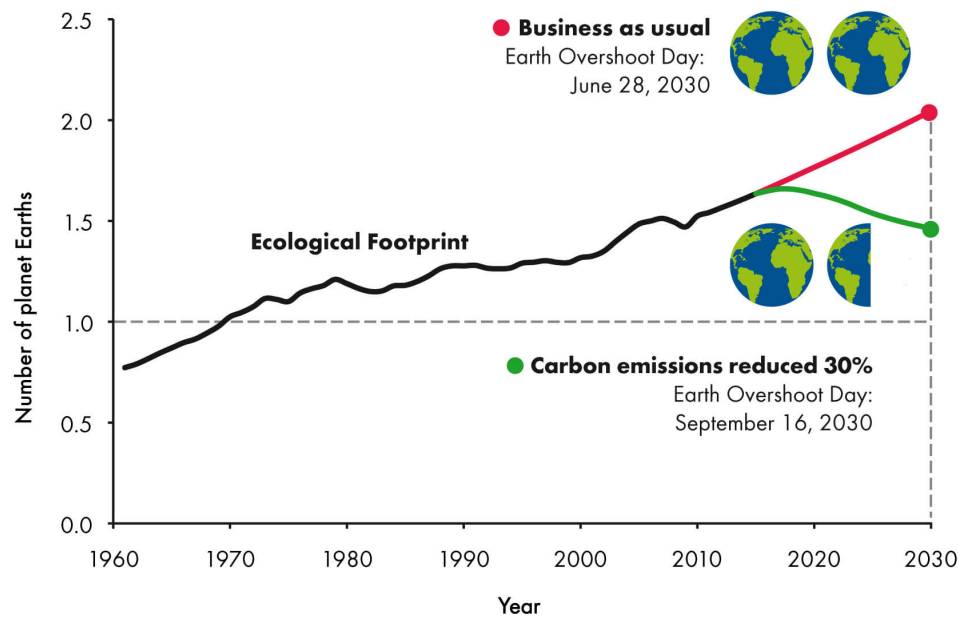


Figure 9.1: The demand for Earths over the years

The overuse of the Earth does not remain without consequences. **The atmosphere is heating up, oceans are overfished, drinking water and fertile soils are becoming scarce. Species are threatened with extinction.** This is also shown by the environmental reports of the UN Environment Programme. Industrial technology and the manufacturing of products are likewise always linked to the extraction and consumption of raw materials from nature and the use of land.

Humanity cannot continue as before; scientists and policymakers worldwide are rethinking their approach. They want to link ecology and social matters with the economy, meaning: a **sustainable development**. For this reason, the 2030 Agenda was adopted in 2015 alongside the Paris Climate Agreement. It offers a great opportunity to achieve the transition toward sustainable and low-emission ways of living and doing business worldwide. The goals of the 2030 Agenda can only be achieved through intensive cooperation within the framework of a new global partnership [31].

9.2 Sustainability at the Company Level

If the description of sustainability is transferred to the company level, the term "sustainability" there describes improving the company's situation today without worsening the **long-term future prospects** for the company, owners, employees, customers and suppliers. To this end, the situation and the consequences of actions must be considered equally across the dimensions of environment, economy, social affairs and technology. The sustainability of a company is reflected in its mission statement (vision, principles), concrete goals and strategies, as well as in the design and characteristics of products, services, business processes, plants and buildings (see Figure 9.2).

The need for sustainable business practices arises from the demands of politics, society, legislation, employees and customers. Companies that fail to meet these requirements may, under certain circumstances, jeopardize their **economic existence and continued survival**. The goals and benefits of sustainable business practices are:

- Preservation of the company's basis of existence and resources,
- Securing and improving economic business success,
- Well-being and satisfaction of stakeholders relevant to the company.

Improving sustainability offers companies a range of opportunities. For example, **higher revenues and competitive advantages** can be generated through sustainable business models, products and services, or cost savings can be achieved through more resource-efficient production processes. Government incentives can, under certain circumstances, also enable tax savings as well as grants or subsidized loans for measures that improve the sustainability of companies [30].

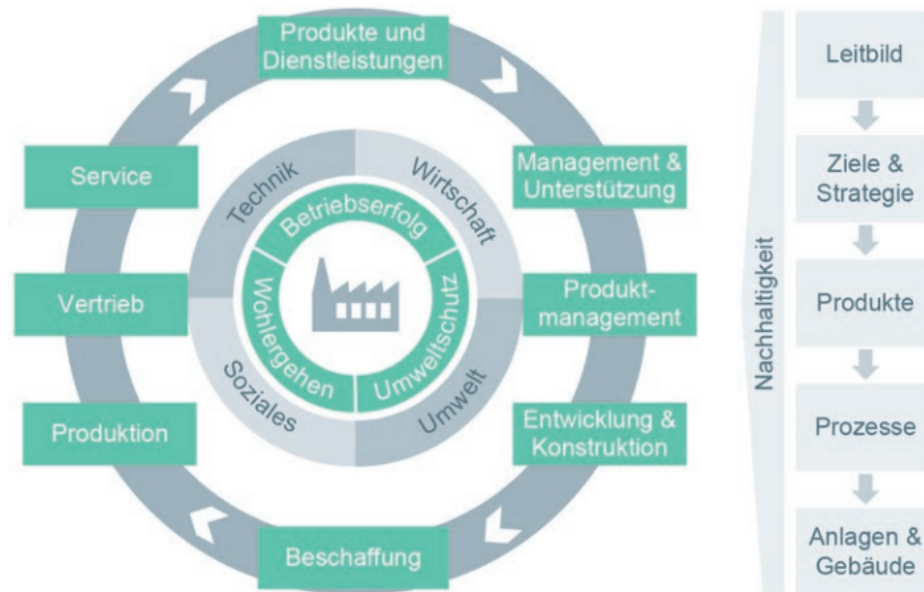


Figure 9.2: Significance of sustainability in companies

9.3 Types of Sustainability

Ecological Sustainability (Environment)

The ecological sustainability of a company is reflected, among other things, in its consumption of energy and materials, as well as in the amount of waste generated and CO₂ emissions. Sustainable companies aim for products and processes that conserve resources as much as possible and avoid harmful effects of their operations on people and nature.

Economic Sustainability (Economy)

Companies can only survive in the market over the long term if they are economically competitive and generate profits. The economic performance of a company is reflected primarily in revenues, balance sheets and productivity indicators. Sustainable companies strive not merely for short-term profit maximization but for lasting economic business success.

Social Sustainability (Social Affairs)

Companies represent a social system of interacting people. For a company to survive, the interests of customers, employees, suppliers and owners must be considered equally. Sustainable companies strive for long-term, trust-based relationships with customers, suppliers and employees, as well as for the protection of the health and livelihood of all parties involved.

Technological Sustainability (Technology)

In an industrial society, the use of technology is indispensable. For companies to survive in the long term, they must keep up with the state of the art. Sustainable companies recognize the opportunities offered by new technologies at an early stage and use them to improve their ecological, economic and social performance [30].

9.4 Opportunities and Benefits of Sustainable Product Policy

Sustainable product policy offers companies a wide range of opportunities and benefits:

- Cost reduction through improved material and energy efficiency
- Adaptation to changing consumer behavior
- Increased attractiveness for skilled workers
- Sustainable products secure the long-term success of the company!

Sustainable Companies Achieve Higher Margins

A sustainable orientation can be advantageous for companies in many ways: it involves greater efficiency, an improved corporate image, employee satisfaction, and also better opportunities in recruiting young talent. In addition, sustainability also has a positive effect on classic economic indicators: sustainably operating companies in the consumer goods and retail sector increase their EBIT margin. On average, it is 6 percentage points higher than that of less sustainably operating competitors.

Sustainability Becomes a Competitive Advantage

One example of this is the company Weleda. For the company, founded in 1921, sustainability has been the top corporate goal from the very beginning. To achieve this, Weleda is also taking new approaches to reporting its economic success. For instance, in its annual report the company no longer looks only at the classic operating result, but also shows in greater detail how this value creation is distributed.

Sustainability – A Trend in Investment as Well

Sustainability is also a major topic among investors, with demand for corresponding investments rising steadily for years. More than half of institutional investors take social and ecological aspects, as well as criteria related to good corporate governance, into account to varying degrees when making investment decisions [32].

9.5 Goals of Sustainable Production

The goal of sustainable production is for the manufacturing of goods to conserve resources while preserving the environment's capacity for regeneration. This secures the natural basis of life for future generations. Environmentally sound production pursues several interdependent individual goals:

- Avoidance of harmful emissions into the environment
- Efficient use of energy and resources
- Transition to renewable energy sources
- Use of sustainably produced resources

- Waste avoidance and the creation of material cycles and recycling
- Restoration and preservation of the biological diversity of ecosystems

Alternatives through Environmental and Efficiency Technologies

Achieving the sustainability goals requires a comprehensive modernization of production systems and of the economy as a whole. New environmental and efficiency technologies play a key role in this. Below are some examples: [33]

- Electric and fuel cell drives
- Electrochemical and thermal energy storage
- Combined heat and power generation
- Bioplastics and composite materials
- Digital networking of systems and supply chains

9.6 Concepts and Methods

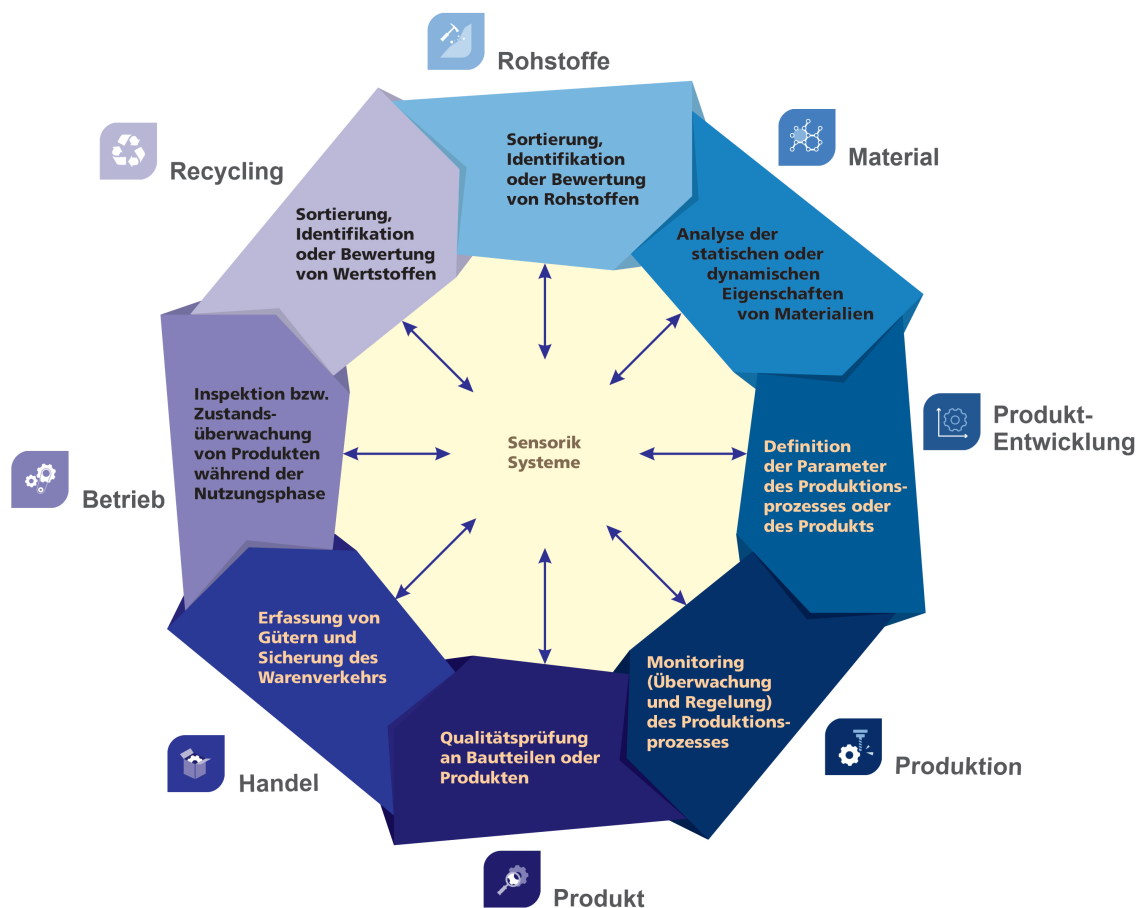


Figure 9.3: Illustration of the product life cycle

To achieve sustainable production, it is not enough to consider only the manufacturing phase of a product. Rather, the **entire life cycle** must be included in the analysis and evaluation (see Figure 9.3).

In both the literature and in practice, various concepts, methods and instruments exist for making the life cycle of goods more environmentally friendly. These mostly aim at improving material and energy efficiency, reducing environmental impact, and creating more sustainable and more responsible supply chains [34].

9.6.1 Sustainable Supply Chain Management

Sustainable Supply Chain Management (SCM) is a strategic approach aimed at promoting ecological, social and economic sustainability across the various stages of a supply chain. The goal is to combine environmentally friendly practices, social responsibility and economic efficiency in order to create a more sustainable and more responsible supply chain (see Figure 9.4).



Figure 9.4: The various influences in sustainable SCM

A central aspect of sustainable SCM is **environmentally friendly procurement**. Companies must ensure that the materials and products they procure are manufactured and delivered in an environmentally sound manner. This can be achieved, for example, by selecting suppliers with environmentally friendly production methods and certifications.

The **reduction of greenhouse gas emissions** along the supply chain is another important goal of sustainable SCM. Companies should take measures to minimize energy consumption and CO2 emissions during the production, transport and storage of goods. This can be achieved, for example, through the use of energy-efficient means of transport, the optimization of routes, and the use of renewable energy sources.

Sustainable SCM also involves promoting social responsibility within the supply chain. Companies should ensure that their suppliers **offer fair working conditions, guarantee the observance of human rights and avoid child labor**. This requires active monitoring of the supply chain and, where necessary, switching to suppliers that are committed to social responsibility. In addition, close collaboration and coordination between the various stakeholders in the supply chain is needed. Companies must work together with suppliers, customers and other relevant partners to define and implement shared sustainability goals. Through collaboration, best practices can be exchanged, resources can be shared, and efficiency gains can be achieved [35].

9.6.2 Circular Economy

The circular economy is a model of production and consumption in which existing materials and products are **shared, leased, reused, repaired, refurbished and recycled** for as long as possible. In this way, the life cycle of products is extended. In practice, this means that waste is reduced to a minimum. Once a product reaches the end of its life, its resources and materials remain in the economy as much as possible. They are thus continuously put to productive use again, in order to continue generating value (see Figure 9.5).



Figure 9.5: Example illustration of a circular economy

The benefits of a circular economy are listed below:

1. Environmental Protection

The reuse and recycling of products slows the consumption of natural resources, reduces the destruction of landscapes and habitats, and helps limit the loss of biodiversity.

Creating more efficient and more sustainable products from the outset would help reduce energy and resource consumption, since it is estimated that more than **80 percent of a product's environmental impact** is determined during the design phase.

A shift toward more reliable products that can be reused, upgraded and repaired would reduce the amount of waste. Packaging is a growing problem, and on average each European generates almost 180 kg of packaging waste per year. The goal is to combat excessive packaging and improve its design in order to promote reuse and recycling.

2. Reducing Dependence on Raw Materials

Resources are becoming scarce. A number of important raw materials are only available in limited quantities, and as the world's population grows, so does demand for raw materials. Finite supplies also mean that some EU member states are dependent on other countries for the supply of raw materials.

Recycling raw materials reduces supply risks such as price fluctuations, availability issues and import dependency. This is particularly true for critical raw materials needed to manufacture technologies that are crucial to achieving climate goals, such as batteries and electric motors.

3. More Jobs and Lower Costs for Consumers

The transition to a more circular economy could boost competitiveness, spur innovation, drive economic growth and create jobs. Redesigning materials and products for circular use would also foster innovation across various economic sectors. Consumers receive more durable and more innovative products that improve quality of life and save money in the long run [36].

9.6.3 Material Flow Management and Analyses

Material flow management is the goal-oriented, responsible, holistic and efficient management of material flows or material systems, with the targets derived from ecological and economic considerations. This achieves:

- Optimization of production processes
- Increase in energy and material efficiency
- Reduction of operating costs
- Avoidance and reduction of emissions and waste

The goals of material flow management lie both in economic profit and in environmental protection. They can be formulated at the national, regional and company level. In what follows, only company-level material flow management will be discussed. There, the goals lie primarily in the **quantification of material and energy flows** and, based on this, in the **optimization of production processes and the increase of energy and material efficiency**. Related goals include reducing operating costs and avoiding or reducing emissions and waste, as well as increasing recycling rates. Within the material flow analysis, these goals are further specified and prioritized within the company. Below, the flow diagram of a material flow analysis is explained step by step:

1. Rough Analysis at the Company Level

The first step usually consists of the rough analysis. It addresses the question of how efficiently energy and materials are used within the company. Using data from all relevant departments, the **input and output quantities** as well as the corresponding costs are determined. This allows the value of the resources used to be directly compared with the resulting products and waste streams. Inputs typically considered here are the **raw, auxiliary and operating materials used, as well as the energy required**. The output comprises the product itself as well as residual materials, scrap and the resulting emissions. Input data can often be determined as a first approximation from controlling, while output data can be determined using production statistics and invoices. Emission data can be determined from exhaust gas or wastewater analyses, and waste heat from the efficiency factor.

2. Defining the Scope of Investigation

Next, the balance boundary of the material flow analysis is defined. For a company-level material flow analysis, not the entire company but only selected process sequences should be considered. The balance boundary of the analysis is narrowed accordingly. The balance period must also be defined. A balance year is often chosen as the period under consideration, since data from the company's balance sheet can be used. The balance period can be further refined down to individual production months or product batches.

3. Naming and Recording of Production Steps

Next comes the recording and naming of the production steps. The production process is broken down into all relevant production steps. Existing process descriptions, equipment registers or cost center plans can be used as a basis. Each production step is then assigned to a representation element (a quantity node) on a diagram.

4. Creating the Flow Diagram

In the next step, a flow diagram is created. In this diagram, the input and output flows for the recorded production steps are captured qualitatively. In a so-called Sankey diagram, material and energy flows can be represented using arrows whose width is proportional to the quantity (see Figure 9.6) [37].

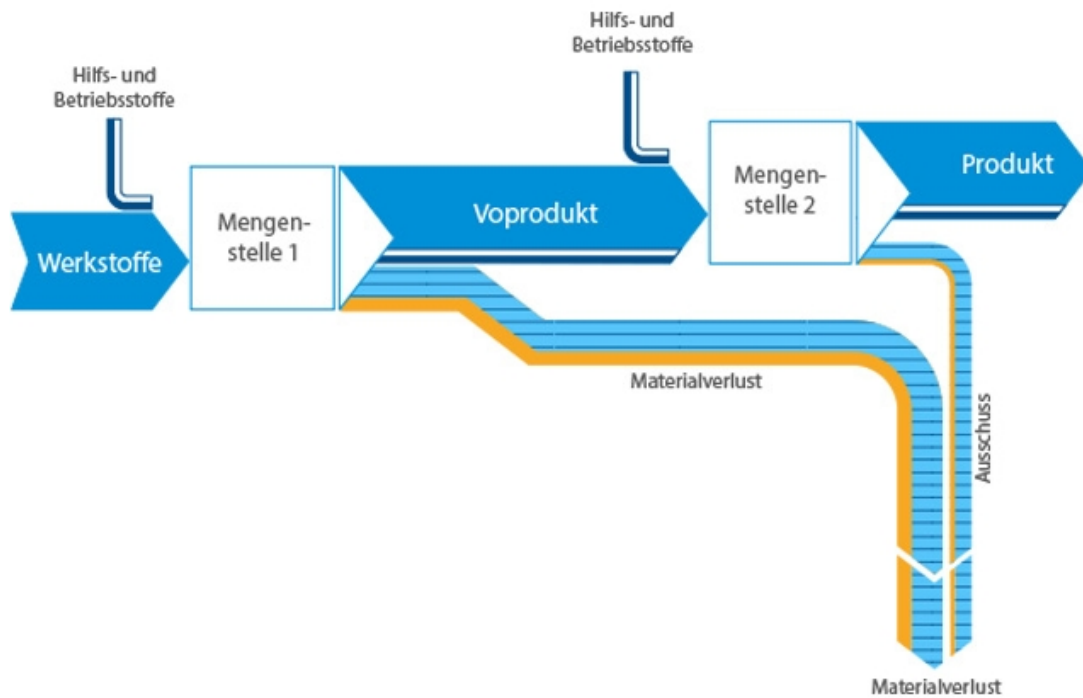


Figure 9.6: Example of a flow diagram in a material flow analysis

9.6.4 Product Carbon Footprint

The Product Carbon Footprint (PCF) is the most established method for determining the climate impact of a product. Throughout a product's entire life cycle - from raw material extraction to recycling or disposal - climate-relevant effects arise in the form of greenhouse gas emissions. The **Product Carbon Footprint** helps to identify and analyze these effects and to reduce or (ideally) avoid them entirely through appropriate measures.

The global climate goals for 2050 aim at a decarbonized world, which is why all products and services must be assessed with regard to their climate impact. This is making the PCF increasingly important.

When calculating the Product Carbon Footprint, a distinction is also made between **direct and indirect emissions** (Scope 1-3): Scope 1 covers all emissions directly caused by the company (combustion of fossil fuels, emissions from chemical and physical processes, etc.). Scope 2 analyzes the indirect emissions arising from external suppliers of electricity, heat, cooling or steam. All other indirect emissions, arising either upstream in the supply chain or during product use, are covered by Scope 3 (e.g. the purchase of products). Figure 9.7 shows the individual scopes and the categories falling under each [33].

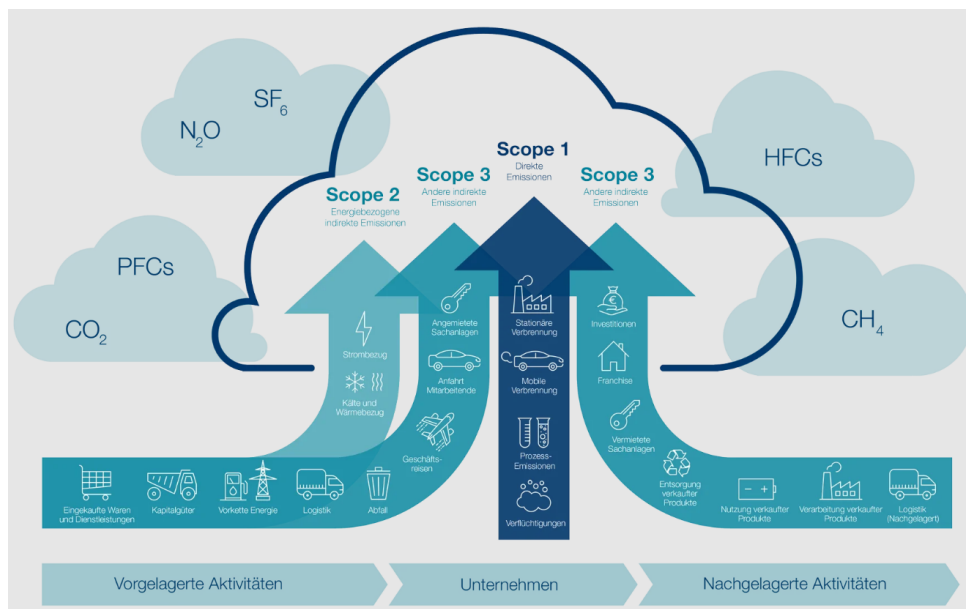


Figure 9.7: Illustration of the individual scopes in the Product Carbon Footprint

9.7 Siemens as a Pioneer of Sustainable Production

Siemens firmly believes that corporate value and the creation of a better future for society can be reconciled. Sustainability holds a **central position across all areas of the company**, from the portfolio to business operations and corporate governance - it is an integral part of the company's identity.

Siemens already possesses innovative technologies and solutions today that can help address the most pressing current and future challenges and support customers and societies in achieving ambitious productivity and sustainability goals. In this context, sustainability is not an add-on to the business but a **fundamental business necessity**. The goal is to help customers and society achieve more with less, and to reinvent people's everyday lives today in order to shape a better world for tomorrow.

These efforts are showing positive results: as early as 2015, Siemens was one of the first major industrial companies to commit to achieving net-zero emissions in its own operations by 2030. In fact, CO₂ emissions have already been reduced by more than half, with the carbon footprint of its own operations having been reduced by around 46 percent since 2019 alone [38].

The following measures were taken:

- Use of green energy
- Installation of solar panels
- Reduction of natural gas consumption
- Implementation of a sustainable fleet policy
- Above all, the use of its own innovative technologies in the field of digitalization and automation

Through such technologies, digital twins for example, production can be made more flexible and quickly adapted to changing requirements and demand. In addition, resource consumption can be minimized by simulating, analyzing and optimizing complex processes in the virtual world.

Flagship Example: Lighthouse Factory Amberg

At its state-of-the-art electronics plant in Amberg, recognized by the World Economic Forum as a Lighthouse Factory, Siemens' manufacturing experts have taken a major step forward. They have developed a digital solution called **SiGREEN** to determine the total carbon footprint of a specific product.

It uses modern cryptographic technology based on a low-energy blockchain to enable the trustworthy exchange of verifiable data, captured at the locations where emissions arise. Interestingly, up to 90 percent of these emissions are not directly attributable to Siemens' own production, but to the **upstream supply chain**.

For this reason, Siemens makes the supply chain an integral part of its decarbonization strategy. The company strengthens local value chains and connects them with globalization, works more closely and purposefully with suppliers, and initiates competition for sustainable solutions within the supply chain. This requires addressing the topic at the point that matters most to every supplier: **at the awarding of contracts**. Only when companies and consumers make more sustainable purchasing decisions do they create incentives for decarbonization in industry.

For these better, more sustainable decisions, SiGREEN provides a reliable data basis for well-founded decisions, specifically for the available products and components. Until now, emissions data along the upstream supply chain has typically been based on industry averages and estimates. Thanks to the Siemens solution, emissions along the chain can be quantified and aggregated into a **dynamic carbon footprint**, while maintaining the data security of the partners involved. This enables more accurate measurement of changes in the supply chain and the definition of more precise reduction measures. Resources can thus be saved efficiently and emissions effectively reduced - in Amberg and far beyond [38].

ESTAINIUM Association

To realize the full impact of this solution, close and trustworthy collaboration with manufacturers, suppliers, customers and partners is required. To this end, Siemens founded the cross-industry ESTAINIUM Association, which drives industrial decarbonization within a pre-competitive



Figure 9.8: Function of the Estainium Network

ecosystem. ESTAINIUM provides a network that enables the **secure exchange and transparent presentation of trustworthy data** on the Product Carbon Footprint (see Figure 9.8). This enables decision-makers within companies to identify where and how they can decarbonize their supply chains. It empowers both end consumers and companies to make smarter purchasing decisions. The priority is to help customers and partners become more productive, more efficient and more sustainable.

DEGREE - The 360-Degree Approach to Environmental, Social and Governance Issues

In the summer of 2021, DEGREE was introduced as a comprehensive framework for sustainability efforts, aimed at making the goals of sustainable and responsible corporate governance more tangible and easier to implement. DEGREE represents a **360-degree approach** to ESG topics (Environmental, Social, Governance) and emphasizes not only the contribution to limiting global warming, but also the consideration of social implications and challenges. Each letter in DEGREE symbolically stands for an area in which Siemens is committed to sustainability: [38]

- D for Decarbonization
- E for Ethics
- G for Governance
- R for Resource Efficiency
- E for Equity
- E for Employability (sustainable procurement capability)

With DEGREE, clear and ambitious goals are set with regard to ESG topics.

Decarbonization

In the context of decarbonization, in addition to the goal of net-zero emissions in its own operations, the company is also aiming for net-zero emissions in the supply chain by 2050. A reduction of 20 percent is targeted by 2030. However, Siemens' main task lies in supporting customers in achieving their own sustainability goals. Renewable energy forms the foundation for sustainable energy generation, while the electrification of existing systems promotes climate-friendly operation across various sectors, particularly in transport through the transition to e-mobility. It should be noted, however, that renewable energy is not universally suited to all industrial activities, such as the need for industrial process heat.

Resource Efficiency

Resource efficiency requires holistic thinking across the entire life cycle of products and manufacturing processes. Companies must move from linear to circular models based on sharing, reuse and recycling of raw materials. Digital technologies such as advanced analytics, 3D printing, the Internet of Things (IoT), machine learning and blockchain play an important role as enablers in complex ecosystems such as the supply chain.

At Siemens, concrete measures are planned, such as increasing the purchase of secondary materials and applying the eco-design approach to all relevant product families by 2030. The company aims to reduce landfill waste by 50 percent by 2025 and to reach zero by 2030. Siemens also provides technologies to ensure reliable water supply and wastewater disposal, to make water management more efficient, and to contribute to the quality and sustainability of water resources, which are of critical importance for the **survival of people and our planet** [38].

9.8 Summary

The topic of sustainability is gaining increasing importance, not only in private households. In order to leave behind a livable Earth for future generations as well, industrial companies must take part in this transformation. Customers and employees now expect companies to act sustainably. And within companies themselves, particularly at the management level, there is a genuine will to operate more sustainably. The fact that an internal sustainability policy brings new opportunities and various benefits should serve as an additional motivation. It is therefore all the more important that policymakers increasingly set requirements in the field of sustainability.

Siemens AG serves as a pioneer in this regard. By implementing a comprehensive sustainability policy or strategy, the company demonstrates how sustainability, held in a central position, can be turned into success while at the same time preventing further damage to planet Earth. Quite the opposite: the environment also benefits from these actions.

Various methods and concepts can be applied to this end. Sustainable Supply Chain Management, the circular economy, material flow analyses and the carbon footprint enable companies both to improve sustainability and to identify potential areas for improvement across the product life cycle.

In addition, the use of innovative digital technologies such as digital twins, advanced analytics, the Internet of Things, artificial intelligence, machine learning and blockchain is indispensable for sustainable transformation in companies.

CHAPTER 10

Smart Logistics

Learning Objectives

After working through this chapter, you will know ...

- ... what Smart, Extra- and Intra-Logistics stand for.
- ... which challenges exist in the field of logistics.
- ... how Predictive and Prescriptive Analytics are applied in the supply chain.
- ... how truck platooning works
- ... why traceability plays a major role in logistics.
- ... how automated guided vehicles are used in intralogistics

Introduction

Digitalization and automation have also had a major impact on the logistics industry. In the broader sense, the term "Logistics 4.0" or "Smart Logistics" describes the effects of Industry 4.0 on all areas of transportation and material supply.

In the narrower sense, Logistics 4.0 refers to the internal and external logistical processes of manufacturing and trading companies and is characterized by their networking, decentralization, automation and standardization. Digital transformation enables higher transparency along the entire supply chain, optimizations with regard to the fundamental logistical requirements (6R)¹, as well as greater flexibility and efficiency [39].



¹ The overarching goal of logistics is to fulfill the 6R: the right product in the right quality at the right time in the right place in the right quantity and at the right cost.

10.1 Future Challenges and Trends

In a study, the German Logistics Association (Bundesvereinigung Logistik, BVL) identified cost pressure, increasing complexity and individualization as current trends in the logistics industry. As in other business areas, digital transformation demands stronger networking and digitalization of business processes as well as greater transparency in value chains from the logistics industry.

Both private and business customers place the highest demands on the **delivery capability, delivery reliability, flexibility** and service level of logistics services. In order to meet increased customer requirements and respond optimally to changing market demands, logistics must **communicate and act faster, more interactively, more reliably and more securely**. This confronts logistics with information-technology challenges. Tracking material, goods and products in real time (**tracking and tracing**) requires the provision of large volumes of data, while at the same time digital services must be integrated into global networks.

Based on these new requirements, new services and business models are emerging, which is why the entire logistics industry must continue to evolve and seize the opportunities of digitalization in order not to become the bottleneck in digital development and Industry 4.0. Information technologies are a decisive driver of this change and form an essential component of new logistics solutions and services:

- **Individualization of logistics solutions and services** for the customer, e.g. user-specific dashboards with logistics KPIs or individually customized apps for shipment management
- **Targeted merging of individual data** to gain new information, e.g. big data monitoring, in order to achieve improved risk management along the value or transport chain as well as better predictability and forecasts (traffic jams, severe weather, delivery bottlenecks, etc.). This enables significant increases in delivery reliability.
- **More efficient service provision** and avoidance of waste, e.g. through cross-functional, IT-supported consolidation¹, automation and risk reduction.
- **Error reduction**, e.g. through the use of sensors, RFID codes, QR codes.

The opportunities for logistics service providers and departments lie in offering solutions and services that create flexibility, reduce complexity, increase efficiency, avoid waste and enable sustainable transport concepts [40].

10.2 Predictive Analytics in Logistics

While Descriptive Analytics deals with the retrospective evaluation of data and thus corresponds to classic reporting, Predictive Analytics attempts to create forecasts and predict future events. The available data is further processed, for example, using data mining or forecasting methods. With the help of mathematical procedures (e.g. machine learning), trends, clusters and preferences are identified and used to generate forecasts.

Predictive Analytics can be applied in many fields, e.g. in production (e.g. machine wear), marketing (e.g. profiling of top customers) and fraud detection (e.g. credit card or insurance fraud).

In the field of logistics, Predictive Analytics is used to analyze flows of goods, information and resources and to enable better monitoring of the value chain. The information gained from

¹ In logistics, consolidation describes the process of reorganizing and bundling products in order to standardize their transport and storage.

big data is used in **Digital Supply Chain Management**¹ to increase efficiency with regard to cost, sustainability and synergies, as well as effectiveness with regard to meeting customer requirements. Predictive Analytics can thus be used to make forecasts about future supply chains and logistical events.

Examples of application areas of Predictive Analytics in logistics:

- **Inventory and shipping planning** can already be prepared months in advance on the basis of purchasing behavior and customer demand, allowing adjustments to be decided at an early stage.
- **Monitoring of logistics performance:** The carrier's transport performance can be monitored in real time in order to determine the exact delivery date. Over a longer period, for example, it is possible to calculate which logistics service provider offers the optimal balance between shipping costs and customer satisfaction.
- **Shipping costs:** Predictive Analytics enables a precise analysis of the causes of rising shipping costs and thus helps to counteract these cost increases.
- **Transport weight:** Costs can be reduced through optimized packaging solutions and weight distribution analyses.
- **Shipping cost transparency:** It can be determined which routes deviate from the calculation more often than average in terms of the costs incurred, or at which points in the supply chain unexpected additional costs arise.
- **Predictive maintenance:** In supply chain logistics, this is used above all in the fleet management of trucks and cargo ships, but also in the areas of order picking and packaging.
- **Sales territory analysis:** By combining cost and customer shipping data, decisions can be made, for example, about a new logistics center in a particular sales territory.

10.3 Prescriptive Analytics in Logistics

Prescriptive Analytics builds on Descriptive Analytics and Predictive Analytics. Big data likewise serves as the basis for Prescriptive Analytics. By examining the effects of different courses of action on the outcome, an attempt is made to achieve an optimal result or the best possible solution. The underlying question behind Prescriptive Analytics is therefore: how must we act in order for a particular outcome to occur? In this way, an optimal course of action can be determined depending on the given situation.

The goal is to **propose recommended actions** and thereby minimize risks or achieve maximum benefits. This analysis method enables **automated decision-making** and reduces the need for human intervention in processes. At the same time, unconsidered influencing factors as well as outdated or limited data can lead to flawed recommendations for action.

In logistics, this analysis method is used for **optimizing delivery routes** and supply chains, as well as in inventory management for the **utilization of warehouse capacities**. It is also referred to as "**Anticipatory Logistics**".

For example, the online retailer Amazon secured a patent for **Anticipatory Shipping** back in 2014. Amazon collects customer data via shopping carts, wish lists, search terms, completed purchases, etc., and uses this to derive customer behavior.

¹ **Digital Supply Chain Management** uses information and communication technology to connect all material, product, personnel and information flows from raw material to end customer across the most diverse logistics channels.

Forecasts for the future are generated from these volumes of data, enabling Amazon to better predict which products the customer will purchase in the near future. This allows the customer's shopping experience to be positively influenced. Overall, the use of Prescriptive Analytics or Anticipatory Logistics makes the logistics industry more robust, more efficient, more profitable and more customer-friendly.

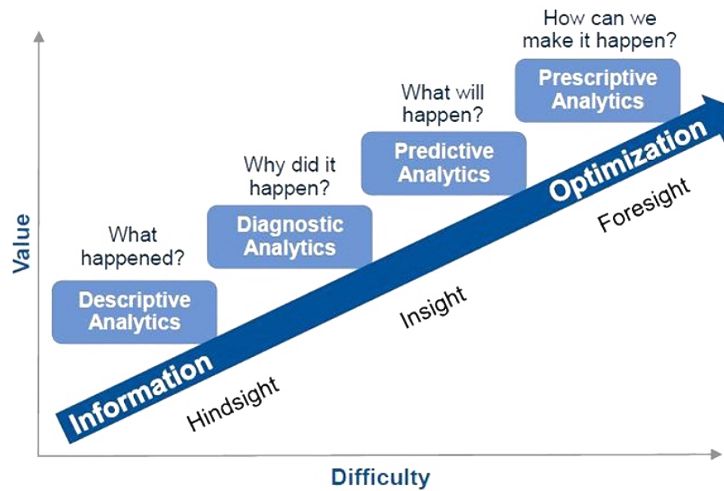


Figure 10.1: The four different business analytics methods according to Gartner (2012)

10.4 Extralogistics

Extralogistics is the antonym of intralogistics and accordingly refers to the logistical flows of materials and goods that take place outside a company's premises, e.g. the transport of goods by a freight forwarder.

Over time, extensive, global transport networks have developed, with the transport of goods largely carried out by truck, rail, ship or aircraft. Corresponding network concepts are increasingly replacing pure point-to-point connections and thereby increasing efficiency. As the complexity of transport networks grows, cost pressure grows with it, while performance promises keep rising.

The transport industry also has to contend with critical external conditions. The International Road Transport Union (IRU), for example, warns of an **acute driver shortage**. Surveys indicate a driver shortage of 21% in the transport sector. At the same time, **growing environmental awareness** is leading to ever stricter regulations. The **penalization¹ of emissions**, e.g. via toll costs, mineral oil tax or emissions taxation of larger companies based on their CO2 footprint, increases the cost per kilometer driven.

In addition, customers place the highest demands on delivery reliability and adherence to individually agreed delivery times. Problems that arise often have to be resolved at short notice, e.g. through special trips or flights. These enormous challenges are accelerating a rethink across the entire industry.

Route / Tour Optimization and Network Planning

Digitalization opens up numerous new possibilities for the logistics industry, including network planning and route/tour optimization in real time. With the help of integrated systems, the transport planner of the future will receive information about the current traffic situation and other relevant conditions.

For example, via a **Supply Chain Event Management (SCEM)** system, they can identify difficulties in the supply chain of their component supplier and, through an early response, prevent problems in their own production, thereby eliminating the need for costly special trips. The software-supported **Transport Management Systems (TMS)** used today face the fundamental problem that they offer no way to connect different areas, such as route and tour planning, with network optimization, in order to master complex customer requirements. In addition to individualizable, cross-functional transport planning capabilities, these solutions also lack the necessary data basis and the intelligence for reliable forecasts.

In order to obtain a comprehensive overview and improved planning options across the entire value network, one needs not only complete transparency of one's own production and logistics, but also information about the production and transport planning of all parties involved (upstream suppliers and logistics service providers) in the upstream supply chain. At present, this visionary approach still lacks an independent, cross-company platform that collects and provides all data via a uniform interface without disclosing trade secrets. Current approaches, however, are striving for the future toward an **intelligent, AI-supported SCEM that enables real-time, scenario-based, proactive interventions** and significantly increases the efficiency of the supply chain.

By recording all shipments seamlessly in order to generate correct master data, TMS can be used to create the foundation for such an integrated overall SCEM system. SCEM represents a further, prospectively necessary stage in the design of logistics. While machine failures can in future be predicted using predictive maintenance, accidents in traffic cannot be planned for. The goal is therefore to include as many external parameters as possible in the SCEM's

¹ (from Latin poena) "punishment"; English penalty

planning system. Through AI, the system can learn and take an increasing number of factors into account.

The **Control Tower** is at the center of the system, as it is the operational monitoring unit of the supply chain from raw material to delivery of goods. At present, it can already visualize all processes along the supply chain via **Track-and-Trace¹ systems**. If, in the future, the Control Tower receives reliable data on the shipment status in real time, this data can be checked for disruptions and the potential scope of the deviation can be analyzed.

Through Prescriptive Analytics, scenarios with cost analyses, delivery time impacts, etc. can be developed and optimal courses of action identified.

In practical projects, the following effects have been observed depending on the scenario:

- System-controlled route/tour optimization: up to 20% efficiency gains
- Network optimization: 25% increase in efficiency
- Meta-optimization between route/tour optimization and network optimization as scenario-based overall control: an additional 8-10% gains and 40% capacity savings in planning
- Application of SCEM: halving of safety stock, reduction of special costs, e.g. for air freight, by up to 70%

At the same time, however, costs also arise for operating the Control Tower and the system. Even the implementation of an integrated overall system takes a great deal of time, money and resources and requires precise planning.

If human influence on the AI is possible, the process must also include approval steps in certain scenarios in order to prevent self-reinforcing effects caused by AI. Thus, despite AI, humans remain the controlling authority and, for the time being, make the final decision [41].

Platooning

Current societal trends such as online retail, rising customer demand and urbanization are leading to a steadily increasing demand for transport. One approach to a solution is the concept of **truck platooning**, which will revolutionize transport on highways. In this concept, several trucks are arranged as a convoy, also called a **"platoon"**, and electronically coupled to one another so that they can communicate in real time.

Technical driver assistance and control systems make it possible to significantly reduce the distance between two vehicles, thereby reducing air resistance. The lead vehicle can transmit its driving behavior, such as braking or accelerating, to the other vehicles, which then act synchronously.

For this purpose, the individual vehicles are equipped with cameras, lidar, radar and steering systems and continuously exchange data via WLAN (ITS G5) (see figure 10.2).

A Cyber-Physical System consists of a part in physical reality and a part in the digital or virtual space:

- In the physical world, the CPS consists of sensors (measuring), actuators (acting), user interfaces, as well as the connecting digital communication technology (communicating).
- At the logical level, intelligent functions handle the acquisition, transport, processing and provision of information (computing + communicating).

¹ Track and Trace: This procedure allows the customer to track their shipment at any time, visually linked to the individual stations along the transport chain. Status recording can be done manually, via app or telematics, as well as using wireless technologies such as Bluetooth or RFID interfaces.

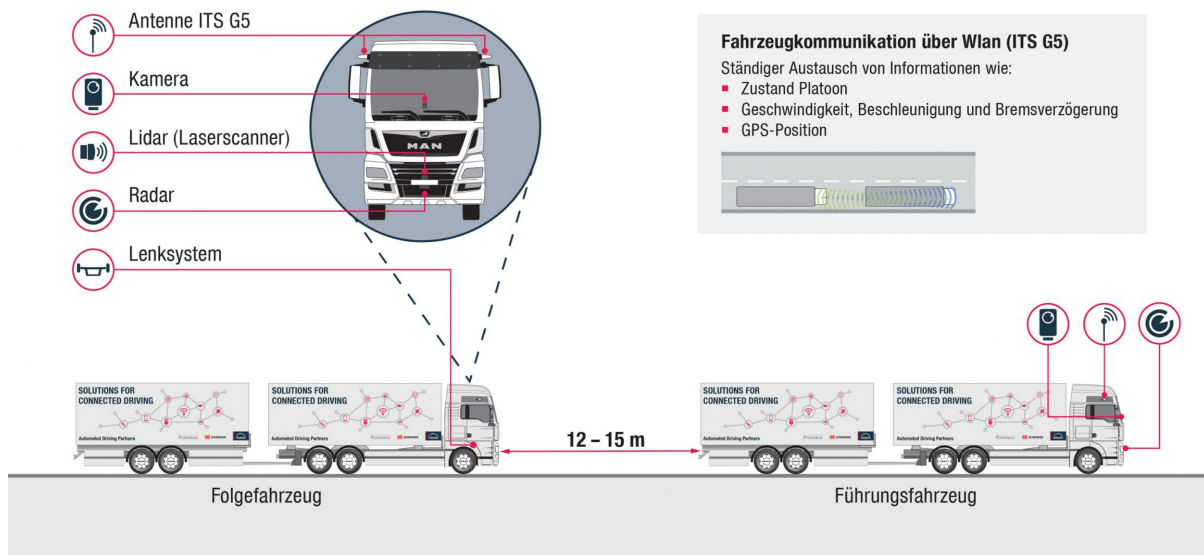


Figure 10.2: Illustration of how truck platooning works

The vehicles can react to traffic situations more predictively, allowing the distance between the trucks to be significantly reduced and thus decisively reducing air resistance as well. Platooning reduces CO₂ emissions and fuel consumption, optimizes traffic flow, makes better use of available road space, and relieves the burden on drivers.

Despite the high degree of automation, the vehicles are staffed with drivers who can intervene at any time and who also take back control when exiting the highway. In the long term, however, platooning is intended to run as autonomously as possible.



Figure 10.3: QR code

Platooning project:
 first transports with
 networked trucks

<https://www.youtube.com/watch?v=dHCUec0EtLo>

Urban Logistics

Urban Logistics refers to the flows of goods within an urban area. It forms the basis for supplying the urban population and accordingly has a major influence on residents' quality of life. Urbanization, e-commerce, the retail sector's desire for more flexible and smaller-scale deliveries, and rising customer demand for more individualized and faster delivery can be identified here as key trends.

As more and more people live in cities while wanting to remain individually mobile, retailers are shifting their ordering patterns toward daily supply of goods while simultaneously expanding their distribution channels, and ever more small-scale shipments are being delivered, traffic volume on the so-called "last mile" is increasing sharply. Solving this problem requires innovative solutions and smart concepts, which can be divided into four basic directions:

- **Vehicle innovations / alternatives:** Emissions can be significantly reduced by switching to alternative drive systems (electric or hydrogen). Deutsche Post, for example, is now increasingly relying on electric vehicles. Cargo bikes as well as autonomous or unmanned delivery options, such as drones or parcel robots, offer only limited transport capacity and are associated with relatively high personnel costs or high acquisition costs.

- **Temporal optimization of traffic load:** By shifting B2B deliveries to low-traffic nighttime hours, transport companies can optimize the utilization of their resources and thereby become more flexible. At the retailer, sales and delivery activities are decoupled, and storage space can be reduced thanks to daily replenishment of goods. Easing time-related delivery restrictions and developing new concepts for low-noise, low-emission nighttime delivery to cities play a decisive role here.
- **Municipal approaches to improving urban infrastructure (city projects):** Municipalities can influence urban traffic volume, ensure higher traffic flow and reduce emissions, not only through local delivery restrictions or the targeted establishment of loading and unloading zones. As infrastructure partners, they can also promote the use of synergy effects between service providers in the form of cooperations. One example of this is the municipally initiated establishment of multi-hubs (large warehouses on the outskirts) and micro-hubs / mobile depots (small warehouses within the city district).
- **Digitalization of the logistics industry:** Digital innovations and automated processes enable optimizations and efficiency gains along the entire supply chain. Customers can choose delivery times more flexibly, which increases the first-time delivery success rate. Optimized route planning enables (cooperative) utilization of transport vehicles and reduces delivery times.

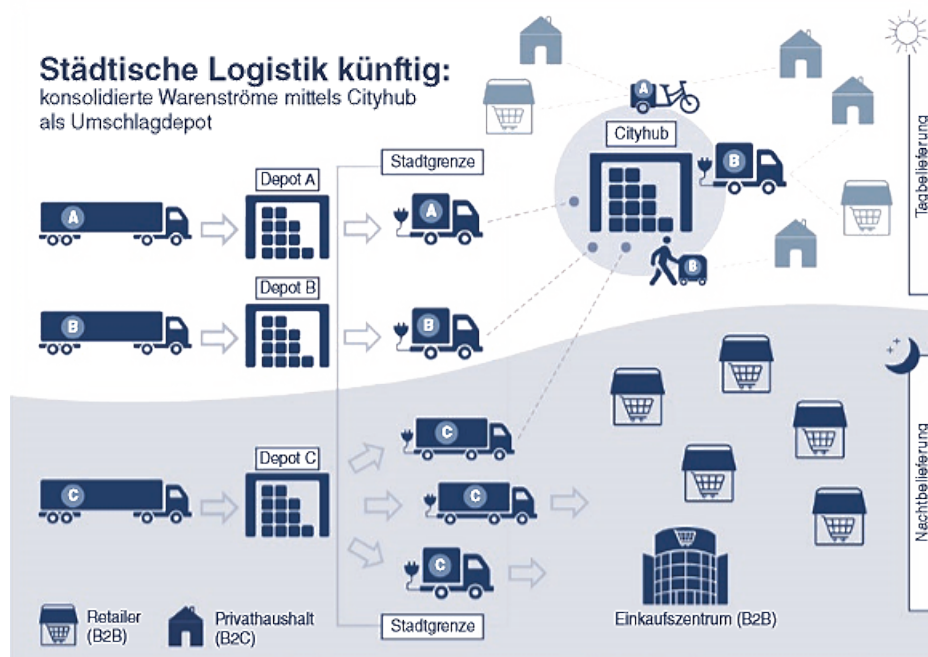


Figure 10.4: Consolidated flows of goods as a feature of future urban logistics

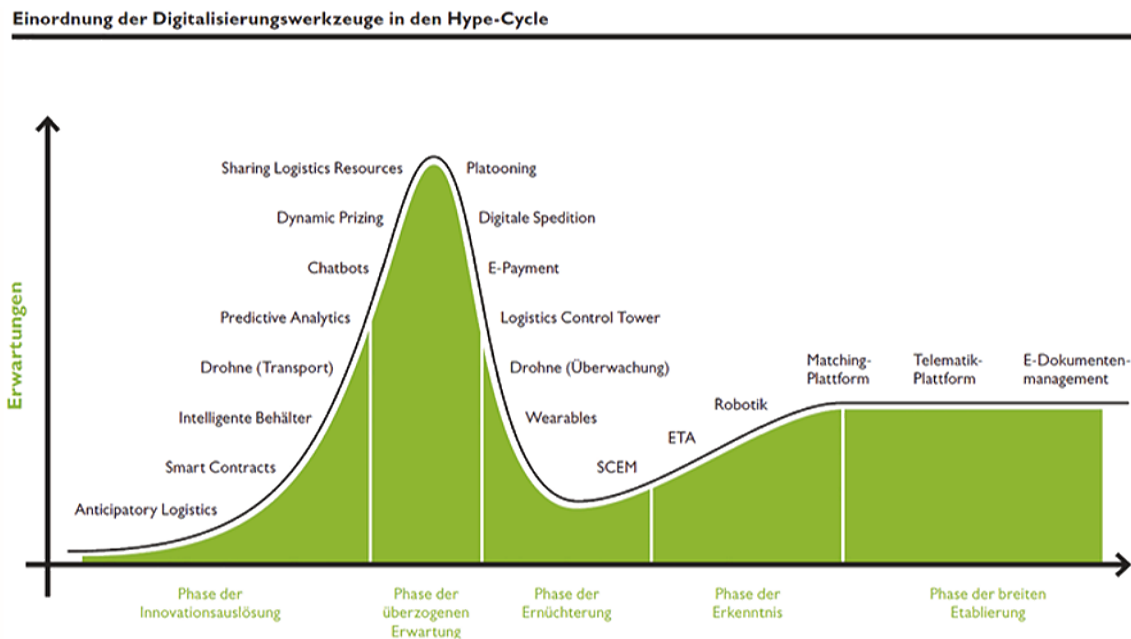


Figure 10.5: Classification of digitalization tools from logistics in the hype cycle
(Source: University of St. Gallen, 2018)

10.5 Intralogistics

According to a definition by the VDMA (German Mechanical and Plant Engineering Industry Association), **Intralogistics** comprises the organization, control, execution and optimization of internal material and information flows as well as the handling of goods within industrial, trading and public sector organizations.

Traceability

The term **Traceability** describes the ability to **capture or use data to reconstruct the history of a shipment**. Today, this is generally done through computer-supported shipment tracking, so-called tracking and tracing. It is used both within and outside the company. Tracking refers to following the raw material through to the end product, while tracing refers to tracing the end product back to the raw material. The combination of both functions results in traceability and enables a particular level of transparency at all stages of the supply chain.

Traceability plays an especially important role in the food industry, ever since EU Regulation 178/2002 came into force in Germany on January 1, 2005. It stipulates the general requirement for traceability of foodstuffs.

In the automotive industry, IATF 16949, discussed in the previous section, likewise requires the marking and traceability of products. This serves to clarify errors after the fact and to limit damage in the event of a defect. In addition, products can be identified at any time and their inspection status determined, thereby avoiding mix-ups.

Unambiguous marking as well as the networking and identification of products in real time is made possible by automatic identification methods. In particular, **RFID technology** (Radio-Frequency Identification) and **barcode technology** are used to track semi-finished and finished parts in production.



Figure 10.6: Marking and traceability along the value creation process

Conveyor Technology and Autonomous Material Flow Systems

Within intralogistics, an efficient material flow through the production systems is essential. To this end, material flow technology within the production system is networked via cyber-physical systems, so that logistics vehicles or individual conveyor modules can communicate and interact autonomously with one another.

In this context, a great deal of research and development is being carried out in the field of plug-and-play conveyor technology. Conveyor technology encompasses all organizational and technical systems for transporting or moving goods and people, generally limited to the field of intralogistics. By transferring the "plug-and-play" principle from computer technology to this field, individual conveyor modules can be added or separated from one another via simple plug connections that contain the necessary interfaces. The following five design rules for plug-and-play conveyor technology have been developed to make permanent and productive use of automated conveyor technology easier for the user:

- **Modularity:** Modules only need to be physically connected to one another; the data connection is established automatically.
- **Function integration:** Each module is equipped with sensors and actuators and has the basic functions of conveying, branching, merging, buffering and handling. Through coordination with one another, the modules can also perform complex intralogistics tasks (e.g. sorting, order picking and sequencing). Identification of the transported load carrier is possible through integration into specialized modules. One current challenge lies in the integration of concepts for personal safety.
- **Decentralized control:** Each module has its own control system. Due to the topology set up by the user, automated topology detection is necessary. A route must be determined for each transport order, and the transport operations are coordinated and carried out by the individual modules. In doing so, deadlocks ("no-way-out situations") caused by modules mutually blocking one another should be avoided.
- **Interaction between modules:** Coordination of transport operations is made possible solely through the interaction between neighboring modules. The goal is to achieve complex system behavior (emergence) through the application of the simplest, most intuitive rules possible, which responds appropriately to the most diverse requirements.

- **Standardized physical and information interfaces:** In order to achieve a simple system design, the modules must have standardized physical interfaces, and the information interfaces must be independent of the modules' specific technical implementation.

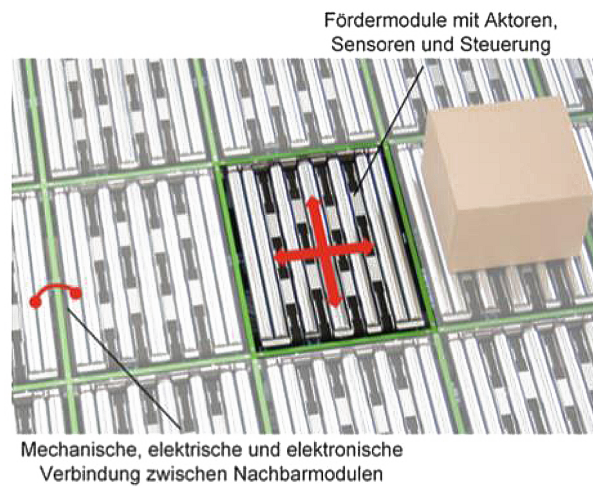


Figure 10.7: Example of plug-and-play conveyor technology modules and connections of the GridSorter

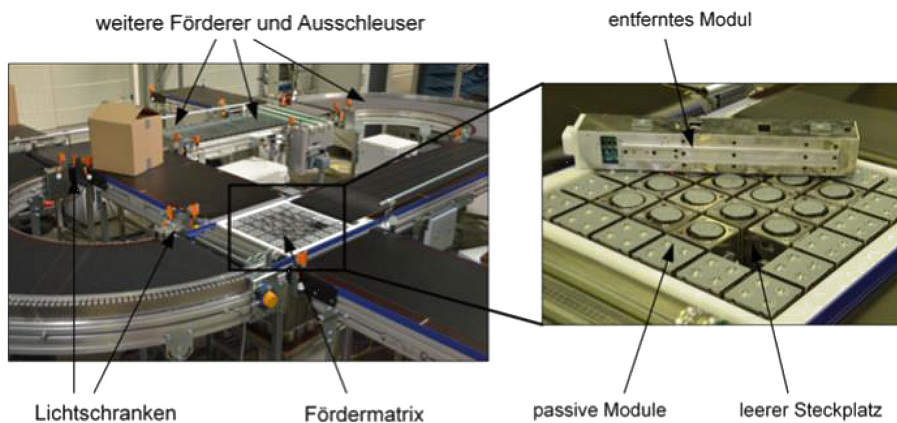


Figure 10.8: Conveyor matrix developed as part of the research project "Cognitive Logistics Networks CogniLog," as an example of plug-and-play conveyor technology

Use of Automated Guided Vehicles

Automated Guided Vehicle Systems (AGVS) (German: Fahrerlose Transportsysteme, FTS) are becoming increasingly important in intralogistics. This is a ground-based conveyor system consisting essentially of one or more contactlessly guided **Automated Guided Vehicles (AGV)** with their own drive, a guidance and navigation system, a data transmission facility, as well as corresponding infrastructure, e.g. for power supply. An AGVS is primarily used for automated material transport (pulling/carrying conveyed goods with active or passive load-handling devices) and can autonomously carry out transport orders along predefined routes.

AGVs come in a wide variety of designs. In general, a distinction is made, depending on the AGV's field of application, between **logistics AGVs** and **assembly AGVs** (for series assembly, where the mount for the assembly object sits on a base structure). AGVs are sometimes also categorized according to the mode of transport and the load being transported, e.g. fork-lift AGVs for pallets or mini-AGVs for small load carriers (KLT).



Figure 10.9: Left: underride AGV with small roller carts for mesh boxes; right: assembly AGV at Volkswagen; in the background, the drivetrain and body are being married on an AGV, while an empty AGV moves away in the foreground

There are three common technologies for powering AGVs: traction batteries (NiCd, lead-acid, lithium-ion batteries), contactless power transmission, or hybrid systems (contactless power transmission plus supercapacitor or battery).

Navigation takes place along a predefined route, either guideline-based via guide wire (inductive), guide lines (optical), guide tape (magnetic) or RFID tags in the floor. Control via navigation systems (laser, GPS, radio) is also possible (see figure 10.10).

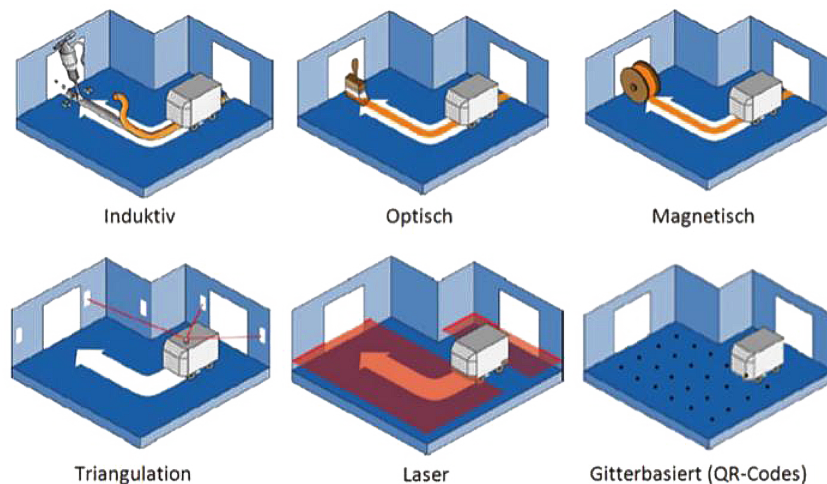


Figure 10.10: Navigation methods for automated guided vehicles

To ensure the necessary safety when using AGVs, numerous laws, regulations, standards and guidelines (e.g. the Machinery Directive, VDI Guideline 2510) must be observed. Among other things, a maximum speed of 1 m/sec (3.6 km/h) is prescribed in areas with pedestrian traffic. In addition, person-detection systems, braking systems, warning devices and emergency stop devices are mandatory for AGVs.



Figure 10.11: a

<https://www.youtube.com/watch?v=B7xmWYOttB8>

<https://www.youtube.com/watch?v=s0re79GRf1Q>

Advantages of an AGVS:

1. Organized material and information flow ensures productivity-enhancing transparency in intralogistics
2. Traceability of logistics processes through a complete process history
3. Reduction of personnel commitment in transport -> lowering of personnel costs
4. High reliability and availability
5. Scalability within the AGVS, adaptation to fluctuating transport demand
6. Simple implementation of intersections and branches
7. Transport of different goods, wide range of additional functions can be realized
8. Only minor infrastructure measures required
9. Minimization of transport damage through high precision in automatic load transfer and pickup
10. Minimization of misdeliveries and waiting stock in the production area through calculable, scheduled and transparent transport operations
11. Generally no additional space requirement, suitable even for low ceiling heights, possibility of layout changes
12. Adaptation to an increasing degree of automation
13. Improved working environment: safe and pleasant working conditions, positive internal and external image, clean and quiet transport operations

Disadvantages of an AGVS:

1. High costs due to the shift of complex technology into the AGV
2. Introducing an AGVS requires precise planning and organization; requirements must be clearly defined
3. AGVS are complex systems; planning, commissioning and, in some cases, maintenance therefore require specially trained personnel
4. Accuracy of GPS navigation is only about 20 m; optical navigation methods are susceptible to contamination; RFID transponders embedded in the floor incur high costs

The largest **user industry** for AGVS is the automotive industry, where AGVs are frequently used directly in assembly or for production supply. The supplier industry has also increasingly relied on AGVS since around 2015. In paper production and processing, paper roll handling is mostly carried out by AGVS applications in order to prevent the damage to rolls that frequently occurs with manual handling. Due to the enormous cost pressure in the industry, AGVS applications are also of interest to the beverage/food industry. For example, Radeberger Brauerei relies on AGVs for pallet handling. Other areas such as the steel industry or hospital logistics are also increasingly being taken over by AGVS applications.

10.6 Summary

The effects of Industry 4.0 on the logistics industry are summarized under the term Smart Logistics. This networking, decentralization, automation and standardization of logistics processes enables greater efficiency, flexibility and transparency along the supply chain.

Current trends such as cost pressure, individualization and increasing complexity are challenging the entire logistics industry. Only if logistics uses IT solutions to act faster, more interactively and more reliably can it meet the growing customer requirements regarding delivery reliability, delivery capability, service level and flexibility.

In Digital Supply Chain Management, Predictive Analytics is used to generate forecasts about future events from big data. These forecasts help, for example, with inventory and shipping planning. Prescriptive Analytics goes a step further and proposes situation-dependent recommendations for action, thereby enabling automated decision-making.

In the field of extralogistics, the entire logistics industry faces major upheavals in the coming years. Digitalization offers extralogistics new opportunities with regard to route/tour optimization and network planning.

Through the use of software-supported Transport Management Systems (TMS), companies can create a data basis in order, in the future, to use an AI-supported Supply Chain Event Management (SCEM) system to enable real-time, scenario-based, proactive interventions in the supply chain and thereby significantly increase supply chain efficiency.

In the field of urban logistics, there are numerous smart solutions that attempt to reduce traffic volume in cities. Despite promising approaches, it remains to be seen which concepts and technologies will prevail in the future.

The field of intralogistics is already characterized today, in many companies, by advancing digitalization.

Traceability plays an important role, as the traceability of products is required across the entire supply chain. Unambiguous marking of products is made possible by automatic identification methods such as RFID technology or barcode technology, which in turn enables computer-supported shipment tracking.

Material flow technology within the production system is increasingly being networked via cyber-physical systems. So-called plug-and-play conveyor technology offers great potential here for autonomous, flexible material transport.

Automated Guided Vehicle Systems (AGVS) are also being used with increasing frequency in many fields, such as the automotive industry or the food sector. Automated Guided Vehicles (AGVs) can be used both in logistics and in assembly, and are available in a wide variety of designs.

CHAPTER 11

IO-Link

Learning Objectives

After working through this chapter, you will know ...

- ... what is meant by IO-Link.
- ... at which levels IO-Link is used.
- ... who administers and drives the technology forward.
- ... which data are exchanged between device and master.
- ... what advantages IO-Link offers.

Introduction

IO-Link is a globally standardized technology for connecting intelligent sensors and actuators in automation systems. The most influential manufacturers in the automation industry have already included a wide variety of IO-Link devices in their product portfolios and continue to develop further IO-Link devices.

The technology is already established on the market and nevertheless offers enormous growth potential. Through the use of IO-Link sensors and actuators, the field level of automation systems becomes intelligent and communicative. It is therefore a key technology for making automation systems, in particular the field level, fit for Industry 4.0.



11.1 Automation Systems and Their Requirements

Automation systems are usually represented in five levels. These five levels form the so-called automation pyramid.

1. Enterprise level (ERP)
2. Operations management level (MES)
3. Process control level (SCADA)
4. Control level (PLC)
5. Field level (sensors and actuators)

The enterprise level represents an entire company's network. Here, so-called Enterprise Resource Planning systems (ERP systems) are used.

Below this, the operations management level forms a network that connects all individual components across an entire plant. Here, for example, production orders are triggered and logistics within production are controlled.

The process control level, one level further down, includes the fieldbus systems. This level serves the user and developer of an automation plant as an interface to the overall system. Here, fault messages and operating states are visualized.

Below that, at the control level, programmable logic controllers (PLCs) operate to evaluate and control all sensors and actuators. The fieldbus systems are also suitable for time-critical applications.

Sensors and actuators are located at the field level. These execute the commands of the PLC, or pass measured process parameters on to the controller. It is at this level that IO-Link finds its application.

Between the levels there are interfaces with different requirements. One of these is the real-time requirement. This describes how quickly data is exchanged between the levels, but also between individual modules. The real-time requirement decreases as the distance of a level from the physical process increases.

At the upper levels, roughly speaking the office workstations, many participants work with large data and data volumes. Here, the processing time of the data is not of decisive importance. At the lower levels, by contrast, the network extent and the number of participants are manageable.

In contrast, small amounts of data, for example process data, are processed here at high speed (microseconds), and due to the direct intervention in the process there is a real-time requirement placed on the system. The highest demands on the communication technology are therefore placed on the interface between the field and control levels. Here, the transmission of process data must take place in real time due to the proximity to the physical process. Specifically, these are the connections between sensors and actuators and the PLC. If signals are sent and received late here, this has an impact on the production process and leads to malfunctions and undefined states.

11.2 Consortium and Brand

„IO-Link is a globally standardized I/O technology (IEC 61131-9) for communicating with sensors and also actuators.“ Behind the technology stands the consortium of the Profibus User Organization e.V. (PNO), which administers and drives the technology forward. Among the more than 200 participating companies, called IO-Link Competence Centers (IOLCC), are, among others, the market leaders of the automation industry and other industry segments such as sensor technology.

The goal of the community is to disseminate the standardized technology and to make sensor and actuator communication flexible and uniform worldwide. Among others, the fieldbus systems PROFIBUS and PROFINET represent such standards. A main goal of IO-Link technology is fieldbus independence. This means that the technology can be used with all common fieldbus systems. Furthermore, IO-Link is specified in the standard IEC 61131-9 under the title Single-drop digital communication interface for small sensors and actuators (SDCI). Another primary goal of the IO-Link community is to establish a quality assurance system that covers the quality of the products, the demo systems, and the service.

However, a manufacturer may only sell under the IO-Link brand if it is an IOLCC or has acquired a license for it. The licensing model, on the other hand, excludes active participation in the community. If a manufacturer holds neither a membership nor a license, it is not permitted to sell under the IO-Link brand. Furthermore, the IO-Link community provides an IODDfinder (Input Output Device Description Finder) on its website. Here, all IO-Link-capable devices from the IOLCCs are listed and uniformly described by the IODDs (Input Output Device Description).

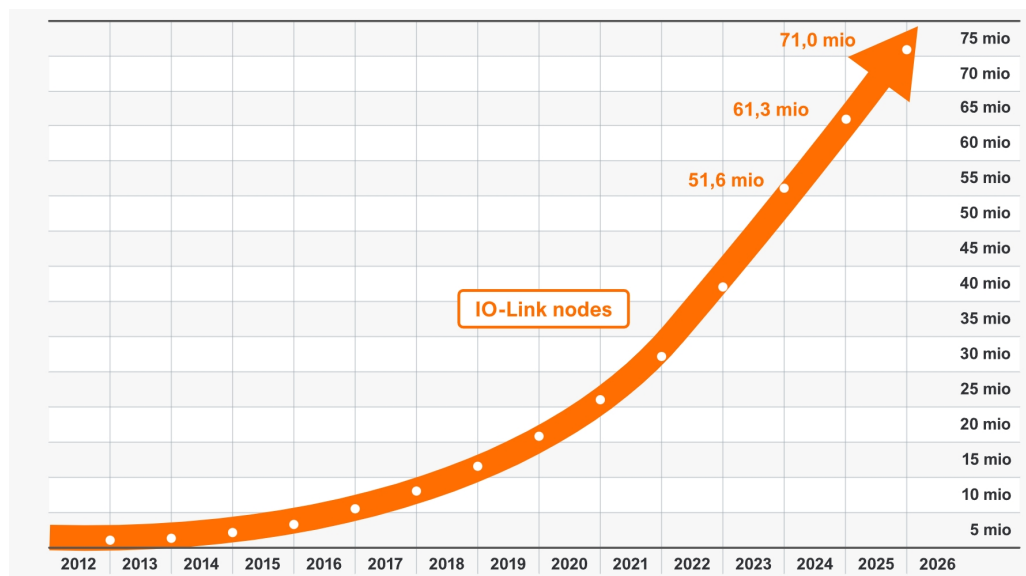


Figure 11.1: IO-Link nodes in the field (2026)

At present, there is no alternative technology competing with IO-Link. Since IO-Link can be represented by any company and is standardized, a competing technology is not necessary. The consortium behind the brand aims to create uniformity, because uniformity improves the performance of IO-Link technology. The efforts made during the development of IO-Link toward the economic success of the technology are already showing effect. As Figure 11.1 shows, by the end of 2025 there were already more than 71 million IO-Link nodes in the field, with the trend continuing to rise.

11.3 Who Manufactures IO-Link Devices and Masters?

IO-Link devices and masters are manufactured by a wide variety of companies. Since all devices are compatible with all masters, there are also companies that manufacture only devices. Most leading companies, however, manufacture both. IO-Link masters differ in the number and type of ports. Typically, a master has between 4 and 10 ports for connecting sensors and actuators. The following companies are known for their active participation and leading position in the IO-Link consortium:

- ifm electronic GmbH
- Siemens AG
- Balluff GmbH
- Pepperl+Fuchs GmbH
- SICK AG
- WAGO Kontakttechnik GmbH & Co. KG
- Festo GmbH & Co. KG
- and many more...

11.4 How Does IO-Link Work?

The standardized communication system IO-Link is a point-to-point communication link between an IO-Link master and a sensor or actuator as an IO-Link device. IO-Link is intended to make the intelligence of the lowest field level of the automation system accessible. The IO-Link master is connected to the fieldbus and, in turn, provides IO-Link ports for communication with IO-Link devices. There are masters for various fieldbus systems, as shown in Fig. 11.2.

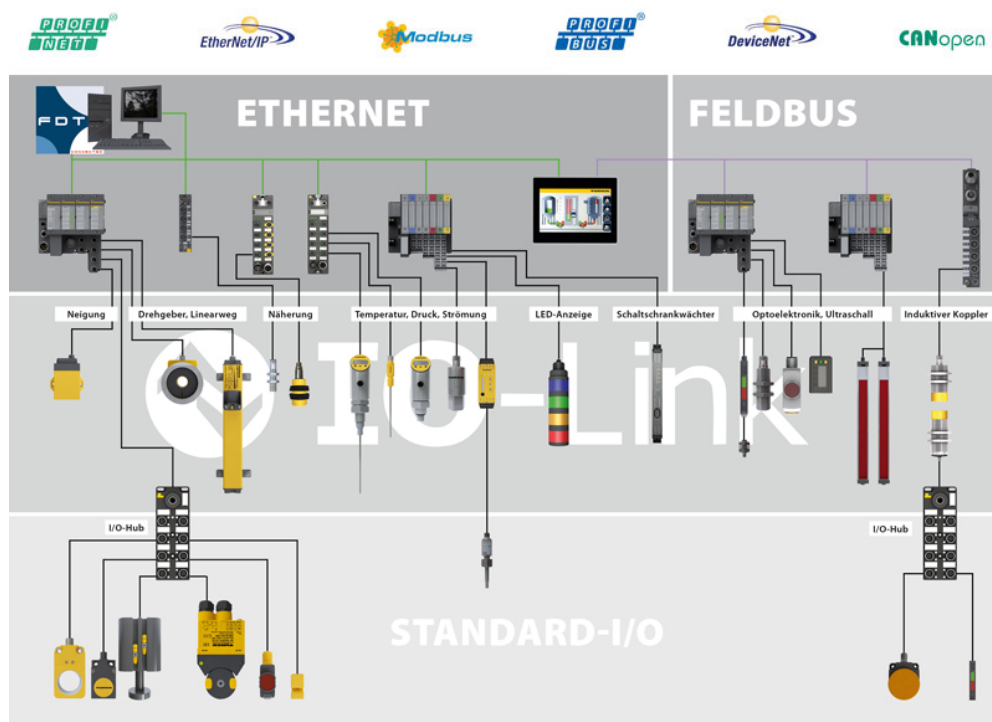


Figure 11.2: Example of the integration of IO-Link into an automation system

IO-Link is not a fieldbus, but a point-to-point communication system. Common fieldbuses have a master that communicates with several subordinate slaves. Since this is not the case with IO-Link, it cannot be described as a fieldbus here. The most common fieldbuses are PROFIBUS, Modbus-RTU, and CC-Link. In addition, there are Industrial Ethernet, which represent a more innovative and faster form of fieldbus. The most common Industrial Ethernet are PROFINET, EtherNet/IP, and EtherCAT. In automation systems, the goal is to achieve continuous process monitoring throughout an automated plant. To achieve this, all layers must be able to communicate with one another. **Operating Modes**

IO-Link fundamentally provides two operating modes, the SIO mode and the COM mode. The SIO mode stands for the Standard Input/Output mode, which guarantees backward compatibility with conventional installations of switching sensors according to IEC 61131-2. In this mode, a binary switching signal is applied to the data line of the IO-Link device. The COM mode stands for the actual communication mode, in which a serial data signal is exchanged between master and device. The following Figure 4 shows both the SIO signal and the COM signal.



Figure 11.3: Operating modes of IO-Link

Transmission Medium

IO-Link distinguishes between two port classes. The standard Port Class A provides for three conductors, which can be supplemented by a fourth. Unshielded cables are used here. The master always has a socket, while the device has a plug. The specification provides for connection via M12, M8, or M5 sockets and plugs, although cabinet components are also available in which the individual wires are clamped. The M-type connection technology achieves the IP65/67 protection class. The established plugs and sockets can be used directly in the field. The following figure shows, by way of example, the pin assignment of Port Class A on an M-type connector.

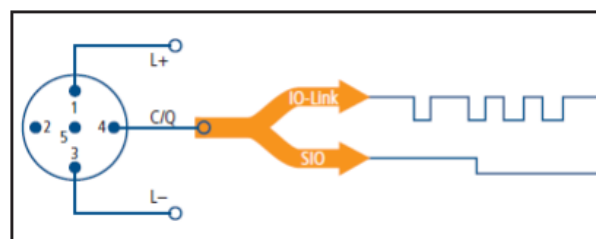


Figure 11.4: Pin assignment of IO-Link Port Class A

For a connection between an IO-Link device and a master using Port Class A, a cable with at least three conductors is therefore required. Via the L+ line, according to the IO-Link specification, a current of at least 200 mA is available at the device. Some devices, often actuators, require an additional power supply. Port Class B is used for this purpose. Port Class B can only be used in conjunction with M12 plugs and sockets, but guarantees a minimum current of 1.6 A. However, the available current is limited to 3.5 A by the M12 connector.

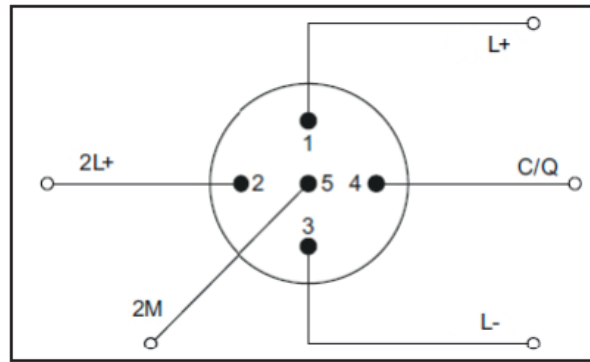


Figure 11.5: Pin assignment of IO-Link Port Class B

Further requirements are placed on the cable between the IO-Link master and device. The maximum cable length must not exceed 20 m, while the conductor cross-section must be at least 0.34 mm². Commonly used sensors such as pressure sensors or position sensors are satisfied by the 200 mA on the L+ line of Port Class A. A commercially available capacitive sensor, on the other hand, has a current draw of around 25 mA, for example.

Data Packets

With IO-Link, data is exchanged between master and device over two logical data channels. Cyclic data is exchanged on one channel, while a second channel handles acyclic messages.

The **cyclic data** contains process data. This is data that is passed directly to the controller of a PLC during operation, for example a digital sensor measurement value. In addition, the process data also includes values that are passed from the controller to an IO-Link device, for example a setpoint for an actuator.

The **acyclic data** is exchanged only on request and essentially consists of parameter and diagnostic data. Parameter data can be used to change a setting in the device, while diagnostic data serve to report certain states in a device.

11.5 IODD

IO-Link devices, i.e. sensors and actuators, enable very complex functions that are not easily handled. For this reason, an Input Output Device Description (IODD) is supplied with every IO-Link device. This IODD contains all the important data concerning the IO-Link device. The IODD is created in the form of one or more XML files. The engineering tools of the controller manufacturers access these files in order to provide the operator with a simple visual representation of the device. In addition to the XML file, images of the device, the manufacturer's logo, and an icon of the device are also required.

The device manufacturer is obliged to create an IODD for every device. The customer acquires this together with the device. The IODD is available for download on the device manufacturer's website. A database of IODDs from well-known IO-Link device manufacturers can be accessed on the internet at <https://ioddfinder.io-link.com>. This website is operated by the IO-Link community. It offers a search engine for IODDs. The user can, for example, search by manufacturer or product designation. However, the IODD finder does not offer a way to compare products on price and technical properties. Alternatively, tools exist that read out the manufacturer ID and device ID of a connected device via the IO-Link master and automatically download the IODD from the website.

The base language of the IODD is English. Only TextIDs are used in the XML file. An IODD can contain further XML files that define TextIDs in other languages. The name of the IODD is strictly prescribed and structured as follows: "CompanyName-DeviceID-Date-IODD1.1.xml". The files are combined into a zip file. The IODD contains the following data:

- **Identification data**

Manufacturer ID: Every IO-Link device manufacturer must be a member of the PNO and is assigned an ID. Device ID: Every IO-Link device is assigned a device ID. A new device ID must be used for every product with a modified function. There is no limit to the number of addressable devices. Because of the wide range of products, market leaders in the industry therefore hold thousands of such device IDs.

- **Parameter data** with address, value ranges, default value, and data type

For each parameter, an ID, an index, the data type, a default value, and the limit values are defined.

- **Process data structure** including length, structure, and value ranges

A specific meaning is assigned to the individual components of the process data; for example, individual bits can be assigned to variables.

- **Event list**

An event is defined for every device-specific event number.

- **Text references** for language files

- **Menu structures**

It is mandatory to structure the IODD according to a specific schema. This is why every IODD must be verified before it is released for use. For this purpose, the IO-Link community provides a piece of software, the IODD checker. This examines the XML file for missing or incorrect content. Upon successful validation of an IODD, the program generates a checksum. The interpreter tool, in turn, calculates a checksum of the XML file the first time it is used. If the two checksums do not match, the tool issues an error message. A modification of the XML file results in a change of the checksum. This procedure prevents the use of subsequently modified IODDs. In this way, the inclusion and use of non-functional IODDs is intended to be prevented.

11.6 Advantages of IO-Link

As a digital interface for connecting sensors and actuators, IO-Link technology offers decisive advantages. A new technology on the market can best establish itself through cost savings for the end user. For this reason, particular attention was paid during the development of IO-Link to economic application in the field. The IO-Link device itself is usually more expensive than a conventional component. The cost savings only materialize during application.

Standardization and Accessibility

One of the main advantages is that IO-Link technology offers an open standard according to IEC 61131-9, where it is standardized under the designation single-drop digital communication interface for small sensors and actuators (SDCI). Among other things, this means that IO-Link devices can be integrated into all common fieldbus systems and automation systems. This is a decisive advantage, since a wide variety of fieldbus systems are used worldwide.

Wiring

Another key advantage is the standardized, plug-and-play style wiring. The simple wiring means that it is nearly impossible to connect an IO-Link sensor incorrectly. The M5, M8, or M12 connections can be attached by hand. No tools are required. Reverse polarity is excluded through the use of M5, M8, or M12 connectors. The unshielded cables are mass-produced items and can therefore be purchased in various lengths at a favorable price.

Remote Parameterization

Since parameter data can be transmitted to the device via IO-Link, on-site configuration is no longer necessary. This brings several economic advantages. On the one hand, the worker in the field only needs to mechanically mount the sensor or actuator and connect it to an IO-Link master. Parameterization can be carried out on the PC. Furthermore, manufacturers offer tools that help the worker set parameters, for example via a visual representation. In addition, automatic data transmission via the fieldbus to the IO-Link master and device is also possible. In this way, data can be automatically provided by a parameterization server. Another advantage is the ability to change settings when retooling a production plant. This is particularly beneficial for hard-to-access units. Remote parameterization is further perfected by IO-Link devices with WLAN interfaces. This allows IO-Link devices to be installed at hard-to-access units without significant cabling effort. Cloud-based solutions are also already being used successfully and are becoming increasingly attractive.

Backup Parameterization

If an IO-Link device in a plant needs to be replaced, the component is automatically parameterized with the data stored in the master. Reconfiguration on site is no longer necessary. Through the identification of the device at the master, an error is reported if an incorrect device is connected to the respective port of the master.

Self-Diagnosis

Self-diagnosis above all results in time savings during the maintenance of an automation plant. Through the self-diagnosis of the devices, the sensor responsible for a production line failure can be quickly identified and replaced. In addition, the devices can report their maintenance status, making maintenance on demand possible. For example, a light sensor reports its degree of contamination, so that maintenance can be carried out at the right time, i.e. during production downtime.

Robust Digital Signal

Due to the high voltage level, signal transmission is very robust and replaces analog signal forms. The continuous quantity is digitized in the sensor and transmitted via the IO-Link interface. As is customary in automation technology, a DC voltage of 24 V is used.

Engineering Tools

In order to minimize the increasing complexity of the devices and their configuration options, the manufacturers of IO-Link components offer a wide range of tools to support the engineering of an automation plant. Tools for parameterization are offered. In addition, the devices are also supported in the programming tools of the controller manufacturers, so that conventional tools can also be used with IO-Link. The tools usually have to be purchased separately and are therefore not automatically accessible to everyone.

Cost-Effectiveness

Most of the advantages just mentioned result in time savings during application. Errors are better detected and found more quickly. This saves money and increases productivity. In an interview with the journal "Elektrotechnik Automatisierung," Jochen Lorenz, the then head of the PNO Steering Committee for IO-Link, reported the following experiences of customers of IO-Link products:

"According to our customers, the use of IO-Link during assembly and commissioning can save up to 30 percent of costs. During ongoing operation, likewise according to customer statements, downtimes can be reduced by up to 40 percent thanks to maximum plant transparency, for example by reading out device information all the way down to the field level."

This once again clearly demonstrates the economic advantages of IO-Link in field application. Another important point for the success of a new technology on the market is investment security for customers. The customer wants to buy a product for which a replacement can still be found in ten years' time. The development of IO-Link aimed to create a standard that offers customers a high degree of investment security.

11.7 Summary

Through standardization, IO-Link technology enables an enormous increase in efficiency in a wide variety of areas. From product development, through the use of the devices, to maintenance, processes are optimized through the standardization of IO-Link. Through the brand's accessibility to companies and the high level of acceptance within the industry, the consortium has created a strong community that continuously applies and further develops its expertise in a targeted manner. The IO-Link brand and technology will continue to grow and play a central role in the context of Industry 4.0.

CHAPTER 12

Predictive Maintenance

Learning Objectives

After completing this chapter, you will know...

- ... what is meant by Predictive Maintenance.
- ... what role Predictive Maintenance plays in the Smart Factory.
- ... what maintenance strategies exist.
- ... what "IIoT" means.
- ... what Remaining Useful Life (RUL) means.

Introduction

Maintenance is a central component of plant management and serves to ensure the functionality of technical systems throughout their entire life cycle. Technical plants are subject to continuous wear and tear during operation, which causes their condition to deteriorate over time. The aim of maintenance is to slow down this process, maintain functionality, and ensure the availability of the plants at the lowest possible cost. In addition to ensuring technical functionality, maintenance thus contributes significantly to the economic efficiency, operational safety, and value retention of production facilities.

Maintenance forms the basis for more advanced, data-based maintenance approaches. Modern strategies build on this fundamental understanding and extend traditional measures to include digital condition monitoring and forecasting methods, which are explained in more detail in the following sections.



12.1 Overview of Maintenance Strategies

To ensure the functionality of technical systems, various maintenance strategies have become established in industrial practice, which differ primarily in terms of the timing and basis of maintenance measures. These strategies form the organizational and technical framework for dealing with wear and tear processes on machines and systems [?].

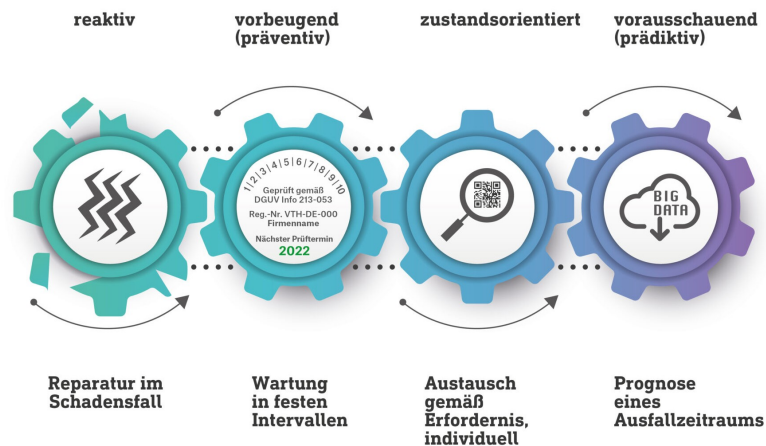


Figure 12.1: Maintenance strategies [?]

Corrective maintenance is a basic form of maintenance. In this case, measures are only carried out after a failure has occurred. The aim is to restore functionality after a malfunction. This approach is simple from an organizational point of view, but can lead to unplanned downtime and high follow-up costs, especially in highly networked production systems [?].

Preventive maintenance, on the other hand, takes a proactive approach. Maintenance measures are carried out at specified time or usage intervals, regardless of the actual condition of the equipment. The aim is to avoid potential failures at an early stage. However, this carries the risk of unnecessary interventions and increased maintenance costs, as the actual state of wear is not always taken into account [?].

Condition-based maintenance is considered a further development of these approaches. Here, measures are planned on the basis of recorded condition data from technical systems. Sensory measurements such as vibrations, temperatures, or operating parameters provide information about the current condition of the plant. Interventions are carried out as soon as defined limit values are reached. This strategy allows for better adaptation to real wear processes and reduces unnecessary maintenance [?].

Predictive maintenance was developed based on this. In addition to current condition information, it also uses forecasting methods to estimate future developments in the condition of the plant. The aim is to determine the optimal maintenance time based on predictions of wear or remaining service life. It thus combines technical condition monitoring with data-driven analysis methods. Predictive maintenance is therefore a further development of the condition-based strategy and forms the transition to modern, digitally supported maintenance concepts [?].

12.2 Definition and Objectives of Predictive Maintenance and Smart Factory

Predictive maintenance describes a data-based maintenance approach in which the future condition of technical systems is predicted so that maintenance measures can be carried out at the optimal time. In contrast to purely reactive or interval-based strategies, this approach is based on the continuous collection and analysis of condition and operating data. The aim is to identify potential failures at an early stage and plan interventions in such a way that unplanned downtime is avoided and resources are used efficiently. Predictive maintenance thus combines condition monitoring, data analysis, and forecasting methods into an integrated maintenance concept [?].

The central objectives of predictive maintenance are to increase plant availability, reduce maintenance costs, and improve the predictability of maintenance measures. By shifting from unplanned repairs to predictable interventions, production interruptions can be minimized and spare parts inventories optimized. At the same time, the condition- and forecast-oriented approach enables better utilization of component lifetimes, as components are no longer replaced exclusively on the basis of fixed intervals. Predictive maintenance thus contributes to increased efficiency, cost-effectiveness, and process reliability in industrial production systems [?].

The implementation of predictive maintenance is closely linked to the development of digitized production environments, which are summarized under the term "smart factory." A smart factory is characterized by networked machines, continuous data collection, and the integration of information and communication technologies into production processes. Within this environment, status data can be continuously collected and evaluated, creating the basis for data-driven maintenance strategies. Predictive maintenance is therefore an essential application of the smart factory, as it uses the availability of digital data to support technical decisions [?].

Current research emphasizes the importance of intelligent sensors and cyber-physical systems for the implementation of predictive maintenance in smart factory environments. The continuous collection and use of status data enables data-based and predictive maintenance. Predictive maintenance can thus be understood as a central use case for Industry 4.0 technologies, in which physical plant conditions are converted into digital information flows and used for operational decisions [?].

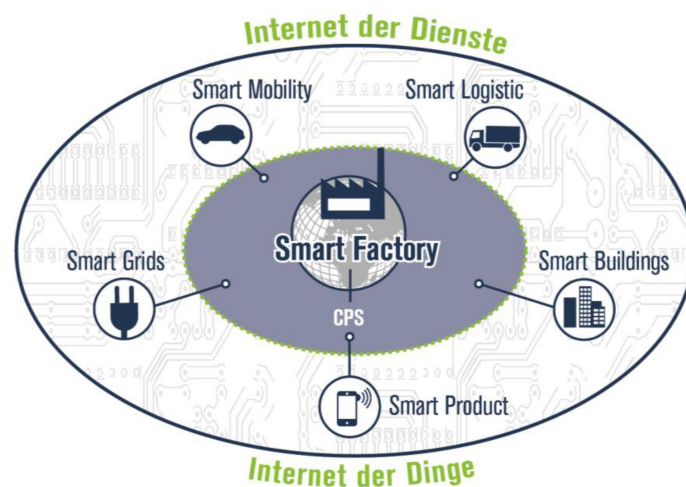


Figure 12.2: SIPM [?]

12.3 Technical Foundations

The implementation of predictive maintenance is based on the availability of technical systems for recording, transmitting, and processing status and operating data. Data-based and predictive maintenance is only possible through the interaction of sensor technology, communication infrastructure, and information processing systems. Technological fundamentals are therefore a prerequisite for continuously monitoring plant conditions and deriving forecasts for future developments [?].

Sensors, Data, and Condition Monitoring

Sensors are the primary source of data for predictive maintenance. They continuously record physical variables such as temperature, vibrations, pressure, or power consumption, thereby enabling an assessment of the current plant condition. The measured values obtained are processed into condition indicators, which allow deviations from normal operation to be detected at an early stage. Continuous condition monitoring thus forms the basis for data-driven maintenance decisions [?].

Industrial Internet of Things (IIoT)

The Industrial Internet of Things (IIoT) describes the networking of industrial plants, sensors, and control systems via digital communication networks with the aim of continuous data collection and availability within industrial processes. In contrast to the general Internet of Things, the focus is on industrial requirements such as reliability, real-time capability, and integration into existing automation structures [?].

In the context of predictive maintenance, the IIoT acts as a connecting infrastructure between the physical plant level and digital analysis processes. Sensor data is transmitted via gateways and industrial communication systems to higher-level platforms, where it is made available for monitoring and further analysis. This creates a continuous data chain that enables continuous condition monitoring and data-based maintenance decisions [?].

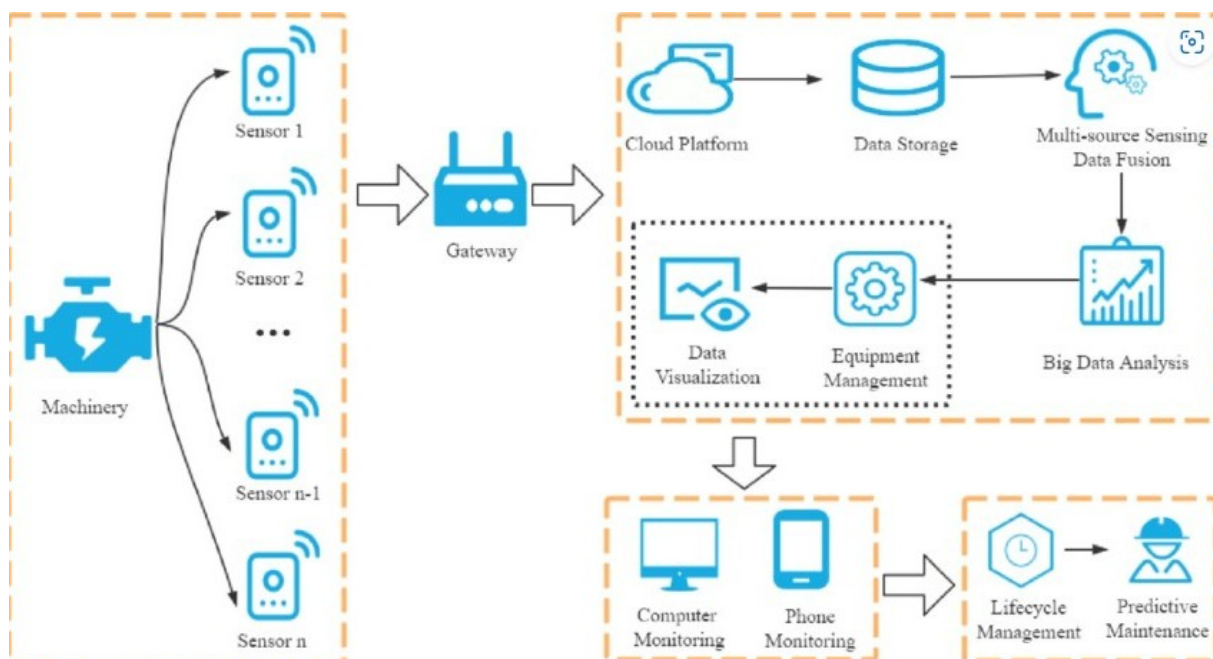


Figure 12.3: IIoT [?]

IT system landscape for maintenance

Higher-level IT systems enable the organizational and operational use of data collected in the IIoT. Central systems include computerized maintenance management systems (CMMS), enterprise resource planning systems (ERP), and supervisory control and data acquisition systems (SCADA).

CMMS support the planning and documentation of maintenance measures, ERP systems integrate maintenance data into business processes, while SCADA systems provide real-time data from production facilities. The interaction of these systems enables the consistent use of data throughout the maintenance process and integrates predictive maintenance into existing production and management structures [?].

12.4 Predictive Maintenance in the Smart Factory

Now that the technological fundamentals have been outlined, the following section focuses on the application of predictive maintenance within networked production systems. The focus is no longer on individual technologies, but rather on their interaction within an integrated production and maintenance system. Predictive maintenance is thus regarded as an organizational and data-driven component of the smart factory [?].

The smart factory as an application environment for predictive maintenance

The smart factory creates the framework conditions for the practical use of predictive maintenance, as machines, sensors, and information systems are continuously connected to each other. This networking enables the continuous availability of status data and forms the basis for data-supported decisions in the maintenance process [?].

In contrast to traditional production environments, information within the smart factory is not processed in isolation, but is used across systems. This allows maintenance measures to be coordinated with production processes at an early stage, reducing unplanned downtime and increasing plant transparency [?].

Integrating Predictive Maintenance into Smart Factory Concepts

Predictive maintenance should not be understood as a standalone function in smart factory concepts, but rather as part of a networked system comprising production, monitoring, planning, and maintenance. Continuous feedback between these areas makes it possible to integrate maintenance decisions directly into operational processes [?].

Monitoring systems continuously provide status information that is used by planning and management systems to efficiently schedule maintenance measures. As a result, maintenance is evolving from a reactive repair function to a strategic process component within the production organization [?].

Data flow and decision support

A key feature of predictive maintenance in the smart factory is the continuous flow of data between the machine level, data platform, and decision support systems. Sensor data is collected via PLCs, transmitted via industrial communication systems, and then stored in databases, where it is available for monitoring and analysis [?].

The processing of this data enables a combination of real-time monitoring and historical analysis. On this basis, early warnings can be generated and potential failures detected, allowing maintenance measures to be prepared before critical conditions arise [?].

12.5 Forecasting Concepts and Remaining Useful Life

Following the integration of predictive maintenance into smart factory environments, attention shifts to the prognostic evaluation of technical systems. The objective is not only to derive the current condition of industrial assets from status data, but also to quantitatively estimate future developments. To this end, various forecasting concepts are applied to assess the remaining useful life of technical components [?].

Definition and significance of remaining useful life

The term remaining useful life (RUL) describes the remaining service life of a technical component until it reaches a defined state of failure or wear. RUL is therefore a prognostic variable that indicates how long a system can still be operated reliably under given operating conditions [?].

Determining the remaining useful life enables proactive planning of maintenance measures. Instead of fixed maintenance intervals, interventions can be carried out based on the actual remaining service life, reducing both unnecessary maintenance and unplanned downtime. RUL is therefore a central component of modern predictive maintenance concepts [?].

In addition, RUL supports decision-making within the smart factory by converting technical status information into an understandable key performance indicator. This allows for better coordination of production planning, resource management, and spare parts procurement [?].

The basic idea of remaining useful life can be represented as a degradation curve with a defined failure threshold (see Figure 3.5).

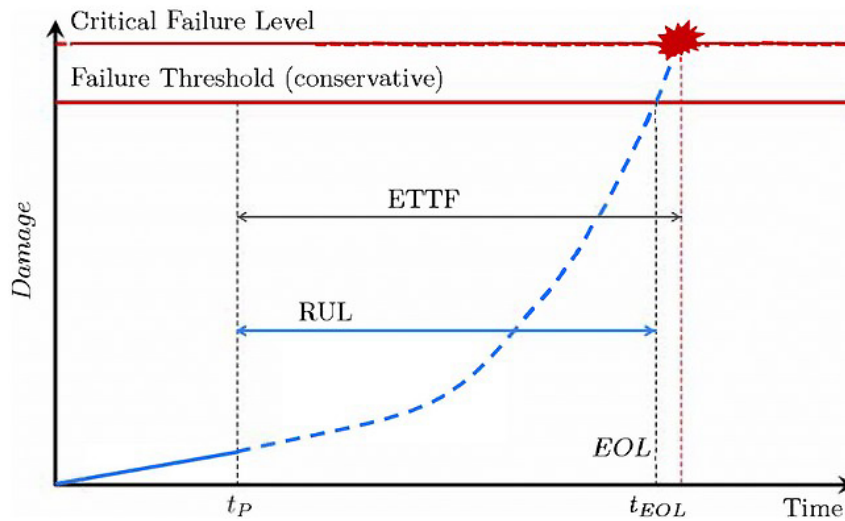


Figure 12.5: Schematic representation of the degradation curve and the Remaining Useful Life (RUL) [?]

Approaches to condition and service life prediction

There are various methodological approaches to predicting remaining service life, which differ in terms of data basis, modeling, and complexity. Basically, three main categories can be distinguished: physics-based, data-driven, and hybrid prediction approaches [?].

Physics-based approaches are based on known wear and aging models of technical systems. Physical relationships are used to mathematically map the degradation process of a component. These methods offer a high degree of interpretability, but require detailed knowledge of the behavior of the systems under consideration [?].

Data-driven approaches, on the other hand, use large amounts of historical and current sensor data to identify patterns and trends. Statistical or learning-based methods are used to make predictions about future state developments. These methods are particularly suitable for complex systems whose physical models are difficult to describe completely [?].

Hybrid approaches combine physical knowledge with data-based methods to leverage the advantages of both approaches. Physical models are supplemented or improved by data-driven analyses, enabling more robust and practical forecasts. Such combined approaches are becoming increasingly important in industrial applications, as they enable both interpretability and high forecast accuracy [?].

Regardless of the approach chosen, the quality of the available data is a decisive factor influencing the accuracy of the forecasts. Reliable data collection and suitable preprocessing steps are therefore fundamental prerequisites for robust life estimation in the context of predictive maintenance [?].

Challenges in predicting remaining useful life

Although forecasting methods are a central element of predictive maintenance, reliably estimating remaining useful life presents a number of challenges. One major difficulty lies in the quality and availability of the underlying data. Incomplete measurement series, sensor errors, or inconsistent data recordings can significantly influence the validity of forecasting models and lead to uncertain predictions [?].

Another influencing factor arises from varying operating conditions of technical equipment. Changes in load profiles, environmental conditions, or usage intensities mean that degradation processes do not always proceed in the same way. Models that have been trained or developed under certain conditions cannot therefore be easily transferred to other applications [?].

Model complexity also poses a challenge. While data-driven approaches can often achieve high forecast accuracy, they require large amounts of high-quality data and increased computing power. Physics-based models, on the other hand, are more interpretable but reach their limits when dealing with complex or only partially known systems [?].

Furthermore, there are uncertainties regarding the timing of forecasts. Particularly in the early life cycle phases of a component, only limited reliable statements can be made about its remaining service life. This requires continuous updating of the models and ongoing evaluation of the forecast quality during operation [?].

Assessment of forecast quality and uncertainties

For the practical application of remaining useful life forecasts, it is not only the prediction itself that is crucial, but also the assessment of its reliability. Forecast models often do not provide exact dates, but rather probability ranges or uncertainty intervals that reflect the possible dispersion of future developments [?].

The accuracy of forecasts is significantly influenced by the quality of the input data, the choice of suitable model parameters, and the timeliness of the available information. Continuous validation of the models using real operating data is therefore necessary to ensure the reliability of the predictions and avoid wrong decisions [?].

In industrial applications, the transparent representation of uncertainties plays an important role in decision support. Maintenance decisions are often based not on a single forecast value, but on an assessment of risks and probabilities. This makes predictive maintenance a supporting tool that relieves decision-makers of the burden of planning maintenance measures without completely automating the final responsibility [?].

Assessing the quality of forecasts is therefore a key component of modern predictive maintenance strategies, as it enables a realistic assessment of the remaining service life while also highlighting the limitations of data-based predictions [?].

12.6 Summary

Maintenance ensures the functionality of technical systems throughout their entire life cycle and makes a significant contribution to availability, economic efficiency, operational safety, and the preservation of asset value. With the increasing complexity of modern production systems, it has evolved from a reactive activity into a strategic corporate function.

In industrial practice, different maintenance strategies can be distinguished. While reactive maintenance is only performed after a failure has occurred and preventive maintenance is based on fixed intervals, condition-based maintenance takes the current condition of the equipment into account. Building on this, predictive maintenance enables forward-looking planning of maintenance measures based on future condition developments.

Predictive maintenance is based on the continuous acquisition and analysis of condition and operating data. Its objective is to detect failures at an early stage, reduce unplanned downtime, and schedule maintenance measures at the optimal time. This approach is closely linked to the Smart Factory, where networked systems, continuous data acquisition, and digital technologies create the basis for data-driven decision-making.

From a technical perspective, predictive maintenance relies on the interaction of sensor technology, IIoT, and higher-level IT systems such as CMMS, ERP, and SCADA. This creates a continuous data flow that enables real-time monitoring, historical analysis, and decision support.

A central element is the prediction of the remaining useful life (RUL). Physics-based, data-driven, and hybrid approaches are used for this purpose. At the same time, RUL prediction is associated with challenges such as data quality, model uncertainty, and varying operating conditions, which makes continuous validation necessary.

In the future, predictive maintenance will gain further importance through artificial intelligence, machine learning, digital twins, and edge computing. In the long term, maintenance is therefore evolving into a strategic optimization instrument within digitalized value chains.

CHAPTER 13

Machine Learning in Industry

Learning Objectives

After working through this chapter, you will know ...

- ... how Machine Learning differs from classical programming.
- ... how the learning process works.
- ... which learning types exist.
- ... which model types are used.
- ... where it is used concretely in industry and where its limits lie.

Introduction

This chapter provides a systematic introduction to the fundamentals of Machine Learning in the context of industrial applications. It first explains what is meant by Machine Learning and how learning-based methods derive models from data. Building on this, central concepts such as learning types, models, and the learning process are introduced. The aim is to convey a fundamental understanding of how Machine Learning can be used to analyze data from the Smart Factory and support industrial processes.

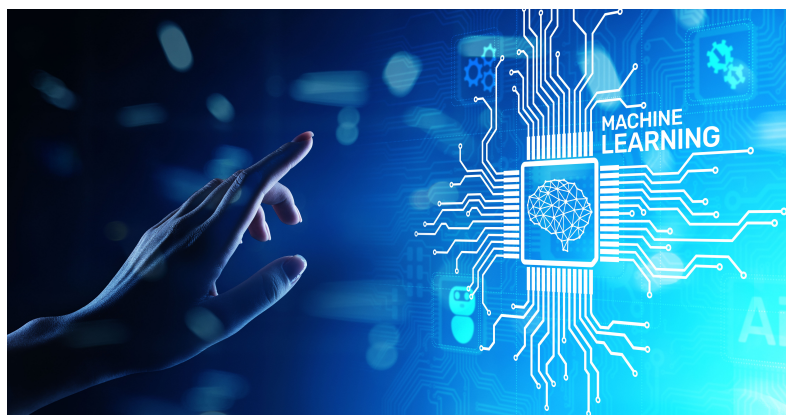


Figure 13.1: Machine Learning

13.1 What is Machine Learning?

Machine Learning refers to an approach in computer science in which computer systems learn from data rather than being controlled exclusively by fixed, pre-programmed rules. The goal is to generate a model from existing example data that describes the relationships between input data and outcomes and that can be applied to new, unknown data. Machine Learning is thus understood as data-based model building, which can be used both to analyze existing data and to predict future states.

In contrast to classical programming, in which all decision rules are explicitly specified, these rules are derived from data in Machine Learning. This means that the system's behavior is not programmed directly but instead emerges through a learning process in which the model is gradually adapted to the observed data. Machine Learning is therefore particularly well suited to applications in which the underlying relationships are too complex or too variable to be fully described in the form of fixed rules.

A simple example from industrial practice is the detection of defective products. Instead of specifying fixed threshold values for measured variables, a Machine Learning system is shown a large number of examples labeled as "defective" or "acceptable". From these examples, the system learns typical patterns and can subsequently evaluate new products automatically. This principle of learning from examples is a central characteristic of Machine Learning methods.

The basic sequence of a Machine Learning system can be divided into three steps:

- Collection and preparation of the data
- Learning a model from this data
- Application of the learned model to new data

Through this approach, Machine Learning systems can also handle large and complex volumes of data, such as those arising in modern Industry 4.0 environments. They are capable of recognizing patterns, making predictions, and supporting decisions without all rules having to be defined in advance.

13.2 The Learning Process

The core of Machine Learning is the learning process, in which a model is generated from existing data that can be applied to new data. The goal is to capture the relationships between input data and outcomes in such a way that the model not only describes known examples but can also generalize to previously unknown data. This process is described in the specialist literature as data-based model building, in which statistical models are learned from example data.

Data Collection and Preparation

The learning process begins with the collection of suitable data. In industrial applications, this data typically originates from sensors, machine controllers, quality measurements, or image processing systems. In order for the data to be used meaningfully, it must be brought into a suitable format. This includes, among other things, removing faulty measurement values, handling missing data, and selecting relevant features. The quality of this preprocessing has a direct impact on the quality of the resulting model.

Model Learning

A model is then learned from the prepared data. In this process, a mathematical model is adjusted so that it describes the observed data as accurately as possible. This fitting process takes place through the minimization of an error, which indicates how strongly the model's predictions deviate from the actual values. The principle of error or loss minimization is a central element of modern Machine Learning methods.

Testing and Generalization

After training, the model is checked against new, previously unknown data. This step is necessary in order to assess whether the model has not merely memorized the training data but is also able to correctly evaluate new situations. A model that delivers very good results on the training data but poor results on new data is considered overfitted. This problem is referred to as overfitting and is one of the central challenges in Machine Learning.

Summary of the Process

The basic sequence of a Machine Learning system can be summarized in three steps:

- Collection and preparation of the data
- Learning a model from this data
- Application of the learned model to new data

13.3 Learning Types

The way in which the parameters of a learning algorithm are determined serves as a criterion for dividing Machine Learning into three fundamental learning types:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

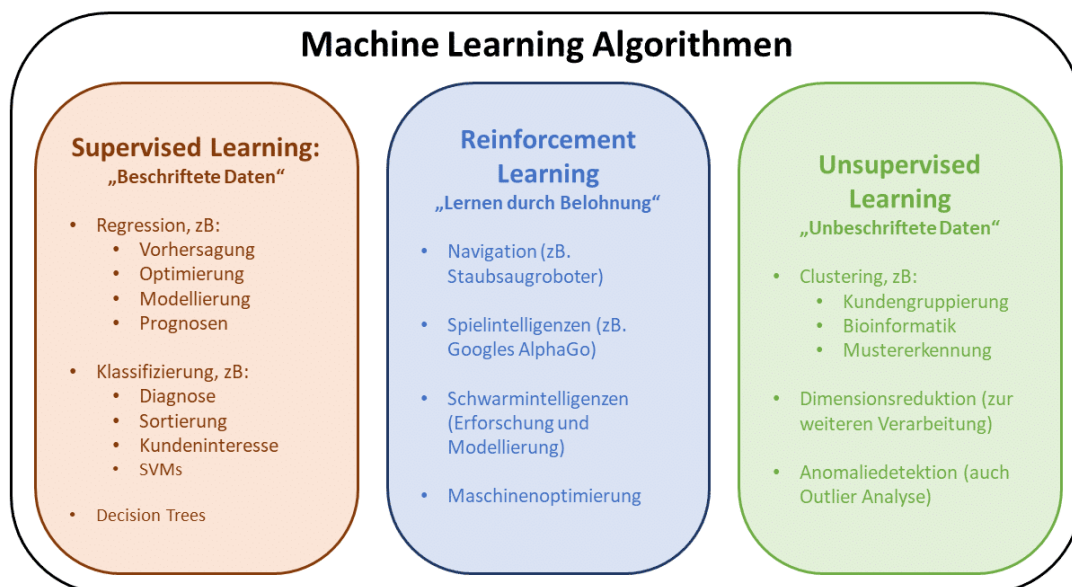


Figure 13.2: Learning Types

Supervised Learning

In Supervised Learning, a dataset is available that is referred to as training data and contains both inputs and the corresponding expected outputs. These expected outputs are also referred to as labels. The goal is to train a Machine Learning model so that it can predict the correct labels when it is later presented with new, similar input data. Learning methods based on such labeled data are referred to as supervised learning.

An illustrative example of Supervised Learning is the automatic recognition of fruit types based on image data. Suppose there is a dataset containing numerous images of apples, pears, and oranges. For each image, the correct fruit type designation is known. These designations constitute the so-called labels.

The Machine Learning model is trained using these labeled images. In doing so, it learns to recognize characteristic features such as shape, color, or surface structure. Once training is complete, the model can analyze a new, previously unknown image and predict, for example, whether it shows an apple.

This exact sequence is illustrated in 13.3: A known image ("This is an apple") serves as a training example. The model learns from such examples and can subsequently correctly classify a new, unknown image.

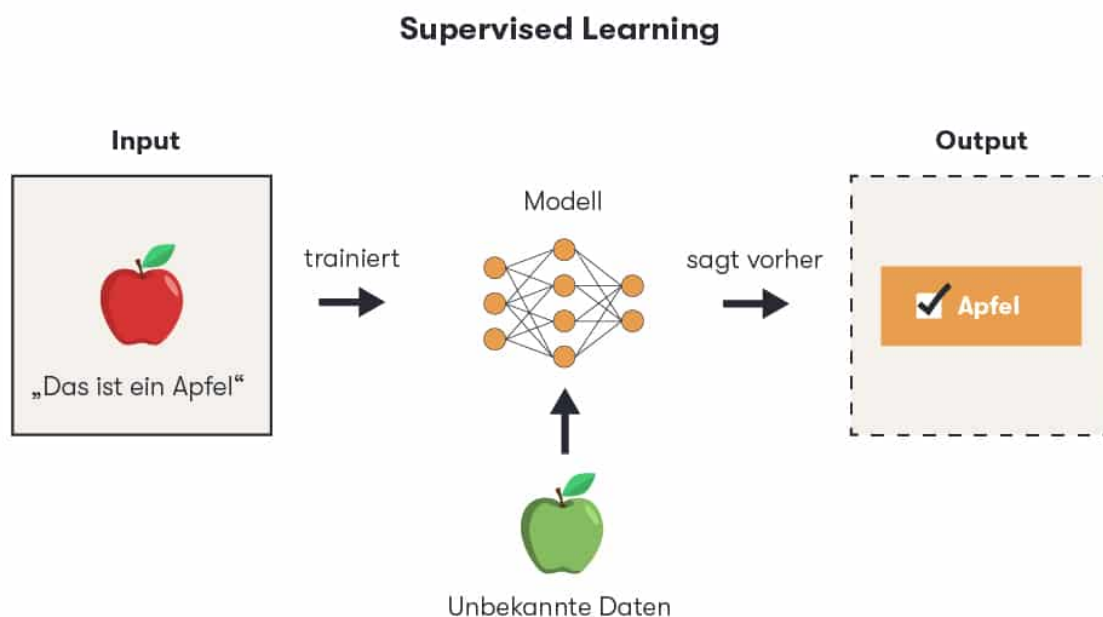


Figure 13.3: Example of Supervised Learning

Unsupervised Learning

In Unsupervised Learning, no labeled data is available. The model receives only input data and attempts to independently identify structures or patterns within it. The goal is to group similar data points or to uncover hidden relationships.

An illustrative example is shown in 13.4: The model is given images of cats and dogs without any designations. It therefore knows no labels such as "cat" or "dog" and receives no information about how many classes exist or which features are decisive. Instead, the model analyzes the image data and extracts typical properties such as ear shape, body proportions, fur structure, or facial contours. Based on these similarities, it automatically sorts the images into groups.

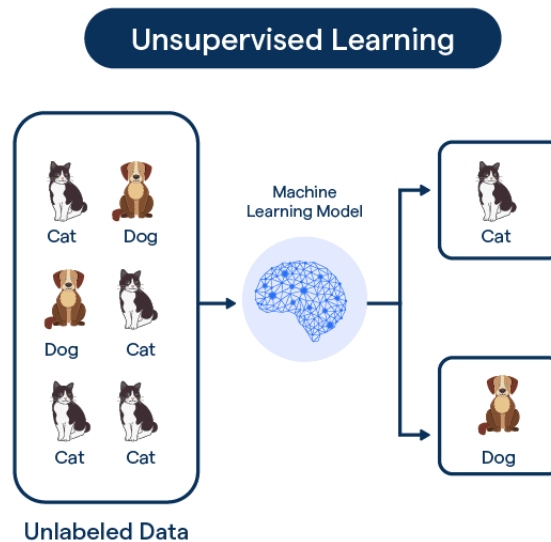


Figure 13.4: Example: Unsupervised Learning

As can be seen in the illustration, two clusters emerge, each with similar features. The model itself does not know that these are "cats" and "dogs" — it merely distinguishes between two groups of image patterns. Only through subsequent interpretation by a human do the clusters acquire meaning. This example illustrates that in unsupervised learning, the structure of the data is identified independently, without target values being defined in advance.

Reinforcement Learning

Reinforcement Learning represents a special form of machine learning that must be considered separately from supervised and unsupervised learning. This learning type is based on feedback from the environment and therefore cannot be classified as fully unsupervised. At the same time, however, no fixed, correctly labeled training data is available, which is why Reinforcement Learning also cannot be classified as supervised learning in the classical sense.

In Reinforcement Learning methods, the system interacts continuously with its environment in order to develop a desired behavior. In doing so, it continuously receives feedback in the form of rewards or penalties, based on which it adapts its behavior. The learning process can be interrupted at any time in order to observe the behavior learned up to that point.

This form of learning bears strong similarities to human learning. From birth, a human baby is already engaged in a continuous process of reinforcement learning by interacting with its environment and learning from the responses it receives.

13.4 Models in Machine Learning

A central element of every Machine Learning system is the model. A model describes the relationship between input data and the corresponding outputs and thus represents the result of the learning process. It can be understood as a mathematical or algorithmic representation that is derived from training data and subsequently used to analyze or predict new data.

Depending on the problem at hand and the data structure, different model types are used. The choice of a suitable model has a major influence on the performance of a Machine Learning

system. A model that is too simple may fail to capture important relationships, while a model that is too complex tends to overfit the training data.

Linear Models

A fundamental model type is linear models. These describe the relationship between inputs and outputs using a linear function. Linear models are comparatively simple, well interpretable, and are frequently used as a baseline, for example for regression tasks or simple classification problems. Despite their simplicity, they are surprisingly powerful in many industrial applications.

Decision Trees

Another important model type is decision trees. They model decision processes in the form of a tree structure in which a decision is made step by step based on conditions. Decision trees are particularly easy to follow and are therefore well suited to applications in which explainability plays an important role, for example in quality control or fault diagnosis.

Neural Networks

Neural networks represent a class of powerful but also more complex models. They consist of multiple layers of artificial neurons that can learn nonlinear relationships between inputs and outputs. Neural networks are used above all when large volumes of data and complex patterns are present, for example in image processing or the analysis of high-dimensional sensor data.

In summary, Machine Learning models differ considerably in terms of their complexity, interpretability, and areas of application. The choice of a suitable model depends on the specific task, the amount of data available, and the requirements regarding accuracy and traceability.

13.5 Data as a Foundation

Data forms the central foundation of every Machine Learning system. Without suitable data, it is not possible to train a model or make reliable predictions. Both the quantity and the quality of the available data have a decisive influence on the performance of a model. In the specialist literature, Machine Learning is therefore described as a data-driven approach in which knowledge is derived from example data.

In industrial applications, data originates from various sources. Sensor data from machines, image data from camera systems, and process and log data from controllers and IT systems are frequently used. This data differs with respect to structure, temporal resolution, and information content. By using different types of data, industrial processes can be analyzed more comprehensively.

Data quality is decisive for the success of Machine Learning methods. Faulty measurement values, missing data, or strong noise can cause a model to learn incorrect relationships. Careful checking and preprocessing of the data is therefore necessary before it is used for training. The literature emphasizes that insufficient data quality cannot be compensated for even by complex models.

In order for data to be used in the learning process, it must be available in a suitable form. The input data is referred to as features, while the corresponding target values in supervised learning are known as labels. The selection of suitable features has a major influence on the performance of a model and can significantly improve accuracy.

To objectively evaluate a model, datasets are frequently divided into training, validation, and test data. The training data is used to learn the model parameters, the validation data to adjust model settings, and the test data for the final evaluation of model performance. This separation is necessary in order to realistically assess a model's ability to generalize.

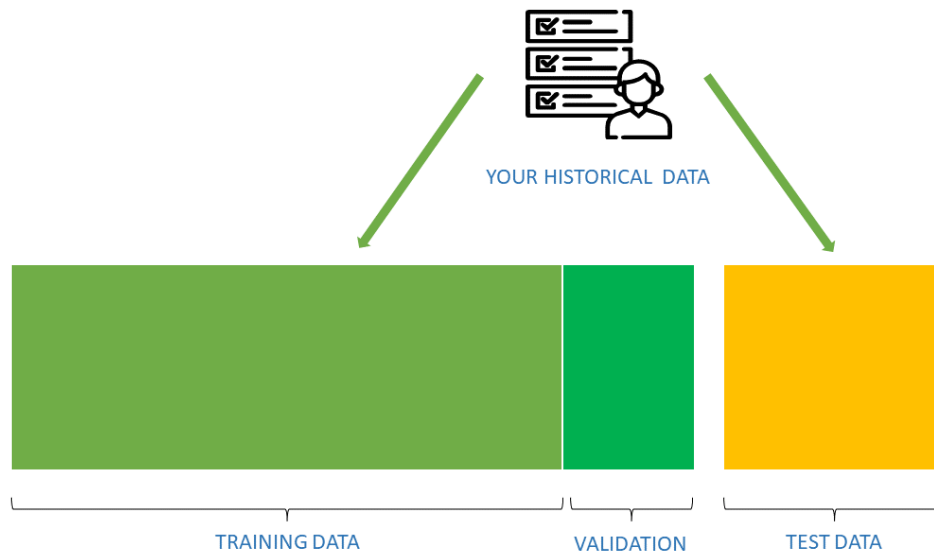


Figure 13.5: Historical Data

13.6 Multimodal Data and Information Fusion

In industrial applications, relevant information frequently originates from multiple data sources. Individual sensor values often provide only part of the overall picture and can also be affected by disturbances or measurement errors. Different data sources are therefore evaluated jointly in order to obtain a more complete and stable picture of the system state.

In a Smart Factory, different types of data arise, for example sensor data such as temperature or vibration, image data from camera systems, and process and log data from controllers. This data complements one another, since it captures different aspects of a process.

Information Fusion describes the combination of such data sources in order to achieve more accurate and more robust results. The data can either be merged before a model is trained or processed separately and subsequently combined.

The use of multimodal data particularly improves the robustness of Machine Learning models. By combining different measured variables, quality problems, process deviations, or unusual machine states can be detected more reliably.

13.7 Applications in Industry

Visual Inspection and Automated Quality Control



Figure 13.6: Visual Inspection in Quality Control

Visual inspection is one of the most widespread applications of Machine Learning in industrial practice. Modern production facilities are frequently equipped with camera systems that continuously capture images of components or surfaces. This image data is evaluated using learning-based models in order to automatically detect defects such as scratches, cracks, shape deviations, or assembly errors. In contrast to classical, rule-based image processing methods, Machine Learning models can capture complex patterns and defect patterns that cannot be clearly defined. This increases both the detection rate and the flexibility when dealing with changing product variants.

A further advantage lies in the consistent inspection quality. While manual inspections can be influenced by experience, attention, or fatigue, ML-based systems deliver reproducible results. Particularly in series production or with high unit volumes, this enables continuous inline inspection without interrupting production. At the same time, detected defects can be statistically evaluated in order to draw conclusions about process causes and initiate quality improvements.

Process Monitoring and Anomaly Detection

In addition to quality control, Machine Learning is used to monitor machines and processes in real time. Models continuously analyze sensor data such as temperature, pressure, vibration, rotational speed, or current draw. The goal is to learn typical patterns of normal operation and to detect deviations at an early stage. Such deviations can indicate incipient wear, faulty settings, or external disturbances.

A particular advantage of learning-based methods is that they can also take into account complex relationships between multiple measured variables. Whereas classical threshold monitoring considers individual parameters in isolation, Machine Learning analyzes multiple signals simultaneously. This allows subtle changes to be detected before visible problems or downtime occur. This form of data-based process monitoring helps to increase equipment availability and reduce unplanned downtime.

Process Optimization and Decision Support

Another key area of application for Machine Learning lies in optimizing industrial processes and supporting operational decisions. Models analyze historical production and process data and identify relationships between input parameters and outcome variables such as quality, energy

consumption, or throughput time. On this basis, predictions can be made for new settings, or optimal parameter combinations can be proposed.

Furthermore, Machine Learning is also used at a higher level, for example in production planning, order prioritization, or demand forecasting. Here, large amounts of operational data are evaluated in order to enable well-founded, data-based decisions. Particularly in complex production systems, this can lead to more stable processes and more efficient use of resources. At the same time, it is important to make the results of such models traceable, so that decision-makers can assess their recommendations from a professional standpoint.

13.8 Limitations and Challenges

Data Dependency and Model Quality

Machine Learning models are strongly dependent on the quality and quantity of the available training data. Incomplete, faulty, or systematically biased data has a direct impact on prediction performance. In industrial applications, additional challenges arise from different data sources, measurement noise, sensor failures, or inconsistent data acquisition. Moreover, for rare fault cases, often only a few labeled examples are available, which makes training supervised methods more difficult. In addition to data quality, model complexity also plays a role: if a model is fitted too closely to the training data, it loses its ability to generalize to new situations. Structured data preparation, suitable validation procedures, and a deliberate choice of model are therefore essential prerequisites for reliable results.

Interpretability and Trust in ML Systems

In addition to accuracy, the traceability of decisions is of great importance in industrial applications. Especially in quality-critical or safety-relevant processes, predictions must be explainable and verifiable. However, many high-performing models, especially deep neural networks, are considered difficult to interpret, since their decision-making processes cannot be readily understood. This can reduce user acceptance and complicate integration into regulated environments. As a result, approaches that increase transparency are increasingly being pursued, for example through explainable models or additional analysis methods for interpreting model decisions. The goal is to strengthen trust in data-based systems and to enable their results to be assessed from a professional standpoint.

Integration, Robustness, and Long-Term Operation

The introduction of Machine Learning into industrial environments is not only a technical but also an organizational challenge. Models must be integrated into existing IT and production systems and work reliably with real-time data. At the same time, industrial processes are subject to constant change, for example due to new products, altered parameters, or wear phenomena. Such changes can cause the data distribution to shift over time and model performance to decline. Continuous monitoring is therefore necessary in order to detect performance degradation at an early stage. In addition, processes for retraining, version management, and maintenance should be defined. Only when the entire model lifecycle is taken into account can a sustainable and stable use of Machine Learning be ensured.

13.9 Summary

Machine Learning is a key technology in the context of Industry 4.0, as it enables the data-driven analysis of complex relationships. In contrast to classical automation systems, which are based on fixed rules, learning-based methods can recognize patterns, make predictions, and support decision-making processes.

This chapter introduced the fundamental concepts of Machine Learning, including the learning process, different learning types, and typical model types such as linear models, decision trees, and neural networks. In addition, the importance of high-quality data and multimodal data sources for industrial applications was highlighted.

Using practical examples, it was shown that Machine Learning is used, among other things, in visual quality control, process monitoring, and process optimization. At the same time, there are challenges, for example with regard to data quality, model interpretability, and integration into existing systems.

Overall, Machine Learning helps to make industrial processes more efficient, more transparent, and more data-driven, making it an important component of modern production systems.

CHAPTER 14

Semiconductor Industry

Learning Objectives

After working through this chapter, you will know ...

- ... why semiconductors are indispensable for Industry 4.0.
- ... how silicon works physically as a semiconductor material.
- ... how semiconductor manufacturing is carried out.
- ... where semiconductors are used in the Smart Factory.
- ... how the global supply chain is structured, along with its risks.

Introduction

Semiconductors are among the central technological building blocks of Industry 4.0 today. Without them, neither the networking of modern production facilities nor the use of artificial intelligence in industrial processes would be conceivable. As the link between the physical and digital worlds, semiconductor devices enable the acquisition, processing, and transmission of data in real time — whether in sensors, processors, or complex control systems.

This chapter examines the role of semiconductors across their entire technological and economic context. First, the physical fundamentals are explained: how does a semiconductor work at the material level, and what properties make silicon the industry's preferred material? Building on this, the manufacturing process is presented, from lithography through etching processes to doping, through which functional components are created from pure silicon.



14.1 Relevance of Semiconductors in Industry 4.0

In the technical world, the use of semiconductors has become an indispensable component of many systems. Semiconductors are the link between physical processes and digital control, i.e., the core of every "cyber-physical system." They form the hardware basis for modern systems such as graphics processors or neuromorphic processors, up to and including artificial intelligence systems. The device and sensor level of Industry 4.0 is based on complex semiconductor systems. These integrate sensing, memory, communication, and logic functions in order to acquire and process data in real time. The individual components of Industry 4.0 require enormous computing capacity and artificial intelligence to analyze large data streams. Semiconductors are particularly important here, as they form the basis for microprocessors, which provide the computing power for large volumes of data. The high relevance of semiconductors for Industry 4.0 can be derived from the rising demand for sensors.

The market for Micro-Electro-Mechanical Systems (MEMS) has been growing continuously since 2010, a trend closely linked to the increasing importance of sensor and actuator manufacturing. MEMS are micro- and nanostructured semiconductor devices that are manufactured using classic semiconductor processes such as lithography, etching, or doping. MEMS are produced on silicon wafers, using process chains that are identical or very similar to those used in Integrated Circuit (IC) manufacturing. It can be inferred that the rising demand for MEMS correlates with an increasing demand for semiconductor manufacturing capacity and technologies. MEMS comprise various sensors, such as oscillators, microphones, pressure sensors, optical MEMS, and radiation sensors. All of these sensors find central application in Industry 4.0 and are responsible for tasks such as real-time condition monitoring (acceleration, pressure, and yaw-rate sensors), machine networking (RF-MEMS), predictive maintenance (microphones, optical MEMS), or autonomous control loops (optical MEMS).

From these developments, it can be concluded that semiconductor devices are indispensable for Industry 4.0. In the course of the increasing networking of production systems and the growing use of artificial intelligence applications and cloud-based services, global demand for semiconductor devices is rising (cf. Fig. 14.1), since these devices form the technological basis for computing power, sensing, and communication.

The following chapter aims to convey why semiconductors are needed, how they are manufactured, what their potential applications are, and how the current supply situation looks.

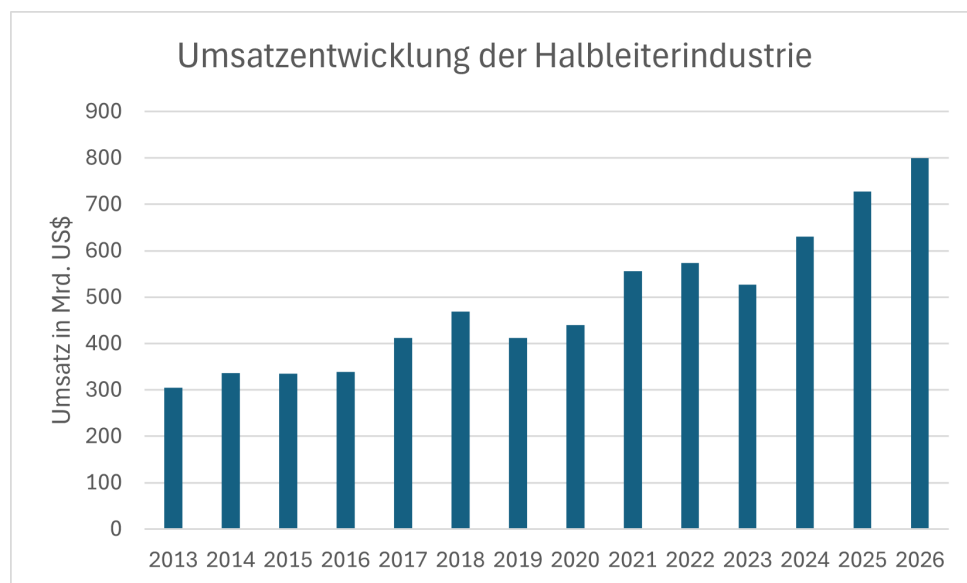


Figure 14.1: Revenue development of the semiconductor industry

14.2 Fundamentals of Semiconductor Technology

Semiconductors occupy an intermediate position between conductors and insulators, with their electrical conductivity being deliberately adjustable. This capability enables controlled signal processing, switching functions, and the storage and transmission of information. In light of this, semiconductors form the basis of all integrated electrical systems and function within Industry 4.0 as the technical link between the physical world and digital control. In order to exploit these properties of semiconductors, the material used in manufacturing is of the highest priority: silicon.

Silicon is the second most abundant element on Earth and an ideal semiconductor material. Silicon is extracted from quartz sand and can be divided into different classes depending on its purity: raw/ferrosilicon has a silicon content of 50–75% and a purity of 98–99%. This type of silicon is used in steel and cast-iron production, whereas products with a higher silicon content are referred to as polysilicon, monocrystalline silicon, or amorphous silicon.

Semiconductor-grade silicon for modern solar cells has a purity of over 99.999999999% (9N–11N, crystalline polysilicon). Monocrystalline silicon is used as the starting material for computer chips, processors, and semiconductors in almost all electronic devices (it consists of a single, continuous silicon crystal without interruptions). It requires an absolute peak purity of 11N or higher. A single foreign atom per billion silicon atoms can already render a microprocessor unusable.

Amorphous silicon has a disordered atomic structure and is used in low-cost solar modules, such as those found in pocket calculators.

Silicon is an element that has four valence electrons in its outermost shell (cf. Fig. 14.2). These enable bonding with neighboring atoms. In the solid state, silicon atoms form a crystal lattice, with each atom forming covalent bonds with four neighboring atoms.

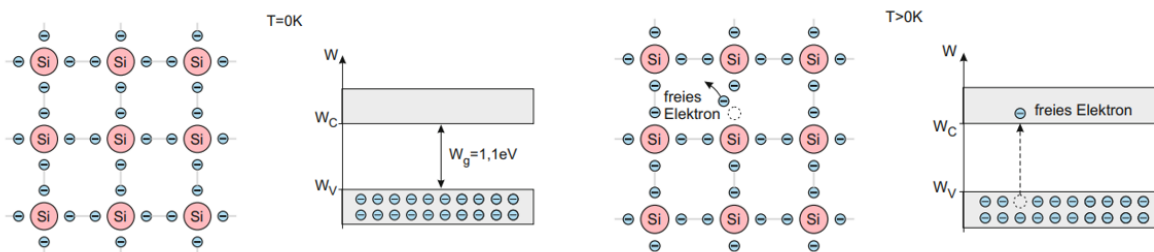


Figure 14.2: Behavior of crystal lattices with increasing temperature

At temperatures near absolute zero, the electrons are tightly bound and no charge exchange takes place. As the temperature rises, electrons can absorb energy. Once the band gap is exceeded, they become mobile. The band gap describes the energy difference between the valence band and the conduction band (cf. Fig. 14.3) of a solid. The band gap of 1.1 eV contributes to silicon not becoming exceptionally conductive at typical operating temperatures. On the other hand, silicon's band gap is ideally suited for constructing field-effect transistors with low threshold voltages. Due to the density of components on processors, it is necessary for the materials used to have good heat dissipation. The use of silicon means that power losses can be dissipated quickly and overheating can be prevented.

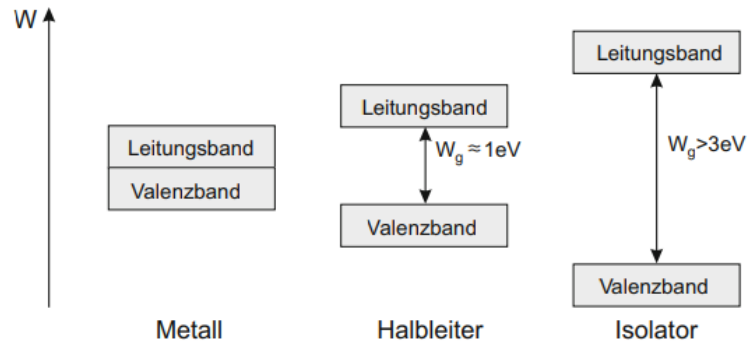


Figure 14.3: Comparison of band gaps (W_G) in electronvolt (eV)

Silicon can be produced as a high-purity single crystal. In this form, the atoms are distributed and aligned absolutely regularly in every direction, without interruptions. This point is essential for the manufacture of microprocessors, since irregularities in the crystal lattice can lead to undesired current paths, which in turn cause short circuits within the component.

Silicon-based semiconductors are not inherently conductive. Conductivity only arises through the deliberate introduction of foreign atoms into the crystal lattice, i.e., doping. Doping is divided into two different variants: n-doping (adding electrons) and p-doping (adding holes).

Doping a semiconductor makes it possible to control its electrical properties. Where the two doped regions meet, a pn-junction is formed, at the contact point of which electrons (n-doped side) diffuse into the holes (p-doped side). Voltage can now be applied in two different directions: the forward direction and the reverse direction.

If voltage is applied in the reverse direction, with + on the n-doped side and – on the p-doped side, the space-charge region widens. As a result, the resistance increases so strongly that only a small current flows, the so-called reverse current (leakage current). Conversely, when voltage is applied in the forward direction, the space-charge region narrows and a current flows.

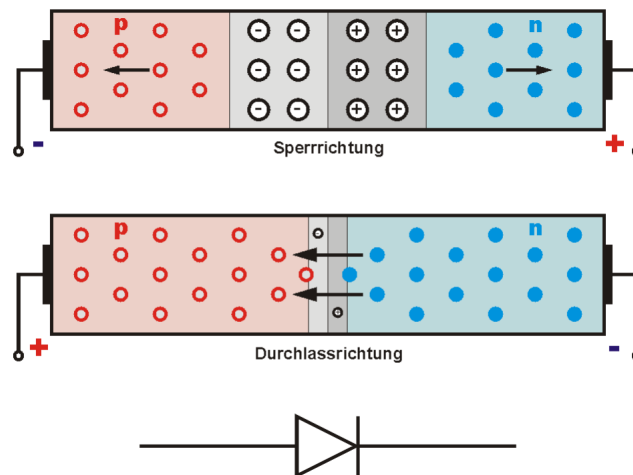


Figure 14.4: PN junction

This effect is applied in a wide variety of components, such as diodes, power-electronic switches, MOSFETs, or IGBTs. An Integrated Circuit (IC) refers to a group of discrete components, which in turn form part of the complex structure of a microprocessor.

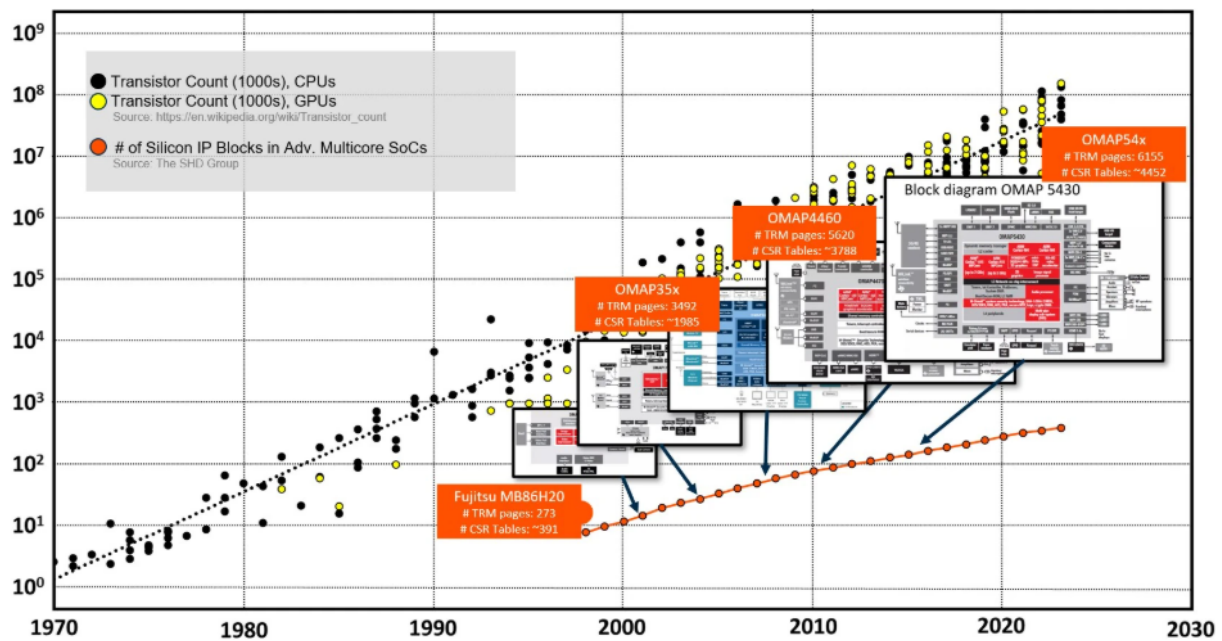


Figure 14.5: Development of transistors on ICs

By integrating a large number of semiconductor-based components, ICs are created that combine complex functions in the smallest of spaces (cf. Fig. 14.5). This development culminates in highly integrated System-on-Chip solutions, in which short signal propagation times and low energy losses are a central prerequisite for real-time-capable industrial applications.

14.3 Lithography as the Key Process in Semiconductor Manufacturing

Lithography forms the basis for the precise structuring of functional semiconductor devices on a global scale. There are various lithography processes — photolithography, EUV, X-ray, electron-beam and ion-beam lithography, and the electron-beam shadow-projection method — which offer different advantages and disadvantages depending on the application. The term ‘lithography’ refers to a sequence of processes that must necessarily be carried out in order to create a functioning basis for ICs. For the sake of simplicity, this chapter deals only with the photolithography process using projection exposure.

The first phase of the process chain consists of preparing the surface of the wafer (cf. Fig. 14.6) by removing moisture from the wafer (1) and applying a functional thin film (SiO_2) (2). This layer is later structured on the wafer. Following this, the wafer is coated with photoresist (3), which is highly light-sensitive and required as a structuring agent for subsequent process steps. It acts as a kind of mask through which a pattern from a design is transferred onto the surface of the wafer.

When the photoresist is applied to the wafer, a small amount of solvent gets onto the surface. To remove the solvent from the applied photoresist, heat must be applied — the so-called pre-bake. Exposure to UV light (4) triggers photochemical processes that cause the solubility of the resist to increase in the exposed areas (positive resist) or decrease (negative resist). This ensures that the desired resist pattern is formed in the subsequent development step, which is then transferred into the underlying layer through etching processes.

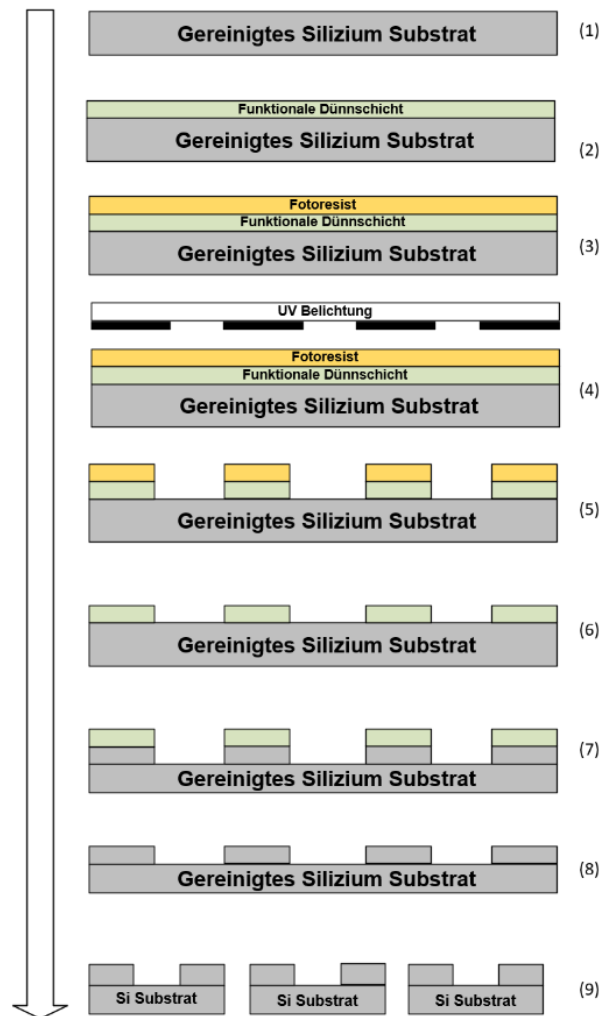


Figure 14.6: Process steps of photolithography

Afterward, the remaining photoresist is removed (6) and transferred into the underlying functional layer or the silicon substrate (7). Finally, the thin film is removed (8) and the microstructures are separated (9). Steps (2)–(8), i.e., the application of new structures, can take place over several cycles, depending on how great the need for microstructuring on the substrate is.

Lithography is the central structuring process of semiconductor manufacturing, as it transfers geometric information into temporary masks and enables the targeted implementation of subsequent process steps.

14.4 Structuring and Doping of Semiconductors

Building on the lithographically defined structure, various etching processes are used to create component structures on ICs in the range of a few micrometers/nanometers.

The etching processes used can be divided into wet-chemical etching methods and dry etching methods. Using these etching methods, the structures and material layers previously defined by photolithography are modified in such a way that the electrical properties of the affected silicon can subsequently be adjusted through doping processes. Following structure definition, selective material removal is carried out so that the structures of the masks are transferred into the underlying functional layers. The removal of the material layers is performed using the two etching methods.

While wet-chemical etching removes material simultaneously in all directions, isotropic structuring achieves uniform material removal in all directions.

The dry etching method aims at exactly the opposite effect and, in particular through plasma-assisted processes, attempts to achieve three-dimensional shaping within the material.

At this point, the various layers on the semiconductor are not only defined through lithographic patterning but are also already physically implemented in the substrate. To ensure that the arrangement of the materials used on the surface of the silicon substrate meets the specifications, insulating, conductive, or semiconducting layers must be applied using thin-film processes.

These layers perform various functions and represent, for example, the gate dielectric or the gate electrode, or the insulation layer of a transistor. Through the interplay of layer deposition and structuring, complex component structures can be produced.

Once preparation of the substrate's material layers is complete, targeted control of the electrical properties of the silicon is required. To achieve this goal, processes such as ion implantation are used. In this process, dopant atoms are introduced into the crystal lattice in order to make the electrical properties of the silicon usable.

These dopant atoms are electrically inactive within the crystal lattice, which leads to changes in charge-carrier concentration, charge-carrier mobility, and specific electrical resistance. These effects are referred to as crystal lattice effects. The dopant atoms are transferred to regular lattice sites through thermal processes and thereby electrically activated. Only through the combination of doping and thermal activation do electrically functional source, drain, and channel regions arise, for example, resulting in a functioning transistor within the silicon substrate.

14.5 Semiconductor Technology in the Smart Factory

The diverse application possibilities of semiconductor components are also found in components of the Smart Factory, where they play an important role. Individual semiconductor devices are not found on their own in the Smart Factory; rather, they occur in combination, for example in processors or sensors, which perform various tasks within the factory. Simulating a wide variety of models, such as the digital twin, requires massive computing power, which is largely generated by processors made up of semiconductor elements.

In principle, semiconductor elements are contained in all networked machines, tools, or other components, which is why, in the course of this chapter, the Smart Factory is divided into three application areas: sensing, actuation, and control.

Sensors are found in both factory automation and process automation and thus contribute significantly to the creation of a fully comprehensive Smart Factory. The sensors installed in factory automation are used to capture geometric quantities and textures of the production environment and enable these textures to be used accordingly by automated/autonomous systems. These sensors also contribute to safety through functional applications such as light barriers. Process automation, on the other hand, is shaped by environmental influences such as temperature, humidity, pressure, or fill level. The capture of these physical quantities by sensors forms the basis for the automation process. The adaptation of processes to the respective environmental conditions is largely enabled by this data acquisition.

Actuators, on the other hand, perform tasks related to adjusting the physical process by converting electrical signals into mechanical actions (cf. Fig. 14.7). One application example of actuators containing semiconductor devices is an electric motor. In an electric motor, energy is controlled

using semiconductor devices such as triacs or thyristors. Semiconductor devices thus also contribute to Industry 4.0 as components of actuators.

Control is the third component of the production of the future and forms the link between sensors and actuators. In order to make concrete decisions regarding the control of machines, a sound data basis is required on which data processing can be carried out. This data processing requires powerful processors, which in turn presuppose the use of semiconductor devices. Signal processors are used for this purpose, which are ICs that perform functions such as filtering, compression, or transformation of the data.

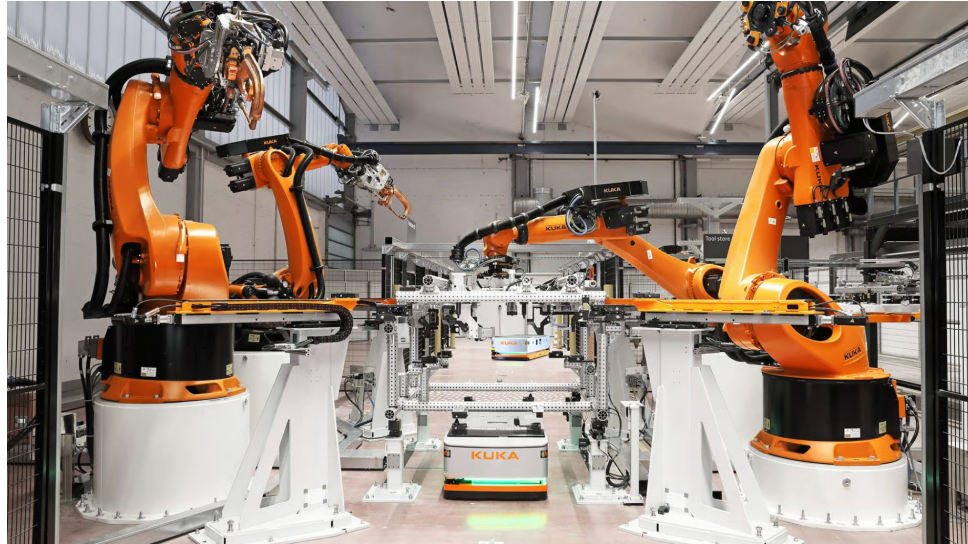


Figure 14.7: Automation using the example of the company KUKA

14.6 Supply Chains, Production, and Logistics of Semiconductors

The majority of the silicon used for semiconductor production is mined in China. The silicon extracted from quartz sand (cf. Fig. 14.8) is subsequently processed by companies such as the German company Wacker Chemie AG to achieve a level of purity sufficient for electrical applications. Following this first production step, the high-purity silicon is further processed into wafers. In this field, the German company Siltronic is the only Western producer of wafers with a diameter of 300 mm.

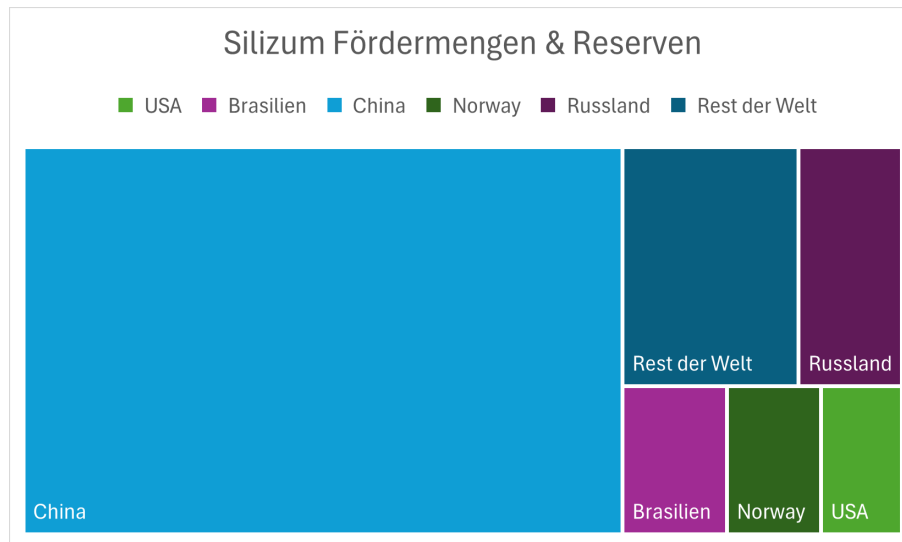
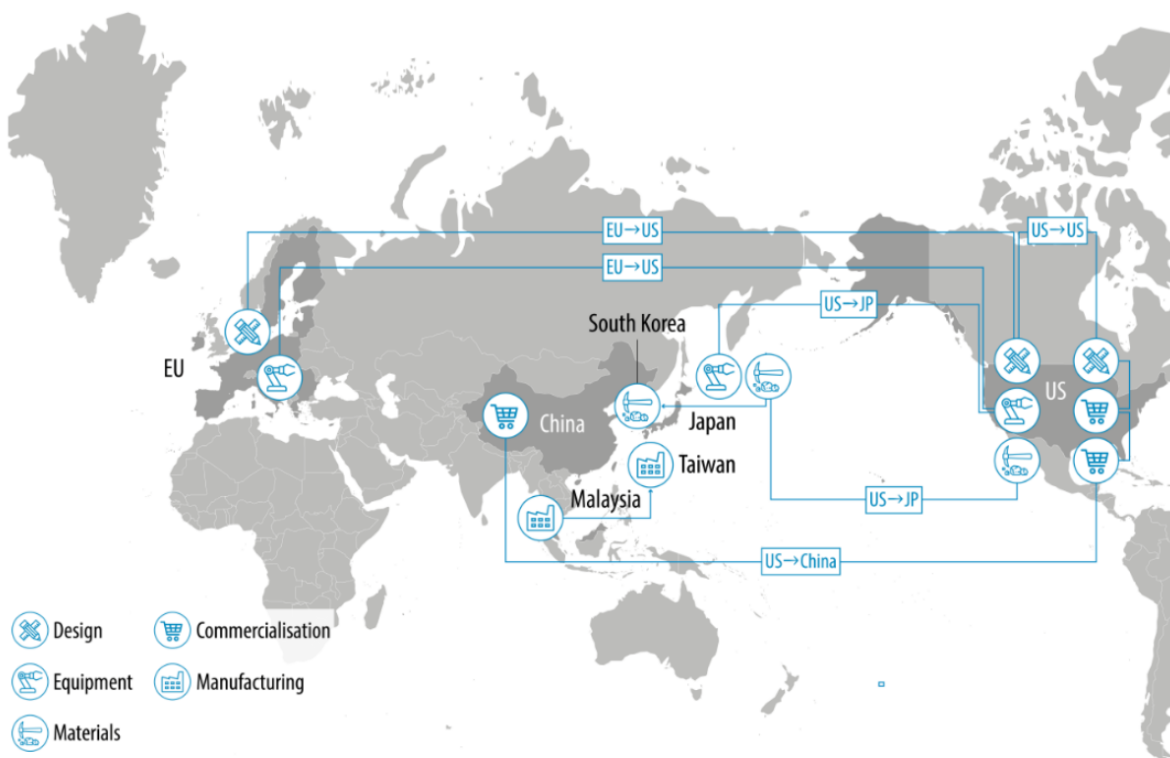


Figure 14.8: Silicon production volumes and reserves

In the subsequent production step, the active components, such as transistors or MEMS, are manufactured in Asia by companies such as TSMC or Samsung using equipment from the Dutch company ASML (cf. Fig. 14.9). In the following process step, the manufactured components are interconnected and process reliability is tested.

To ensure quality, electrical tests are performed, and to prevent subsequent damage, the finished semiconductor parts on the circuit board are covered with a package. Finally, the finished semiconductor device is shipped to the customer, who integrates it into their product and delivers it to the end customer via various channels.



Source: EPRS.

Figure 14.9: Global interconnection of semiconductor supply chains

14.7 Challenges and Trends

Demand for semiconductor components has fluctuated sharply for years. In the course of the global pandemic caused by the coronavirus (SARS-CoV-2), there was a significant increase in demand for semiconductor chips. This increase was so substantial that, in some cases, the supply of chips could no longer be guaranteed, or only at a price level that exceeded the financial resources of many households. This resulted in production stoppages and short-time work among the workforces of German automotive manufacturers such as Audi, BMW, Daimler, and Volkswagen. These events illustrate the significant dependence of the German automotive industry on foreign semiconductor suppliers.

The observed drop in supply is attributable to a series of events that affected the semiconductor industry in various parts of the world. These include, in particular, fires at several semiconductor manufacturers in Asia as well as a winter storm in Texas that brought production in that state to a standstill. This crisis was not only the result of short-term events but was also due to the structure of the semiconductor industry itself. Because 15% of the chips used in premium vehicles are already outdated by the time they are installed, procuring them proves difficult.

A small number of semiconductor manufacturers, including Nexperia, Infineon, and Renesas, focus on the automotive market. These manufacturers are also dependent on raw materials from Chinese production. In contrast, renowned manufacturers such as TSMC or Samsung primarily serve the demand of companies such as NVIDIA.

Semiconductor demand vs. supply (Strategy& model)

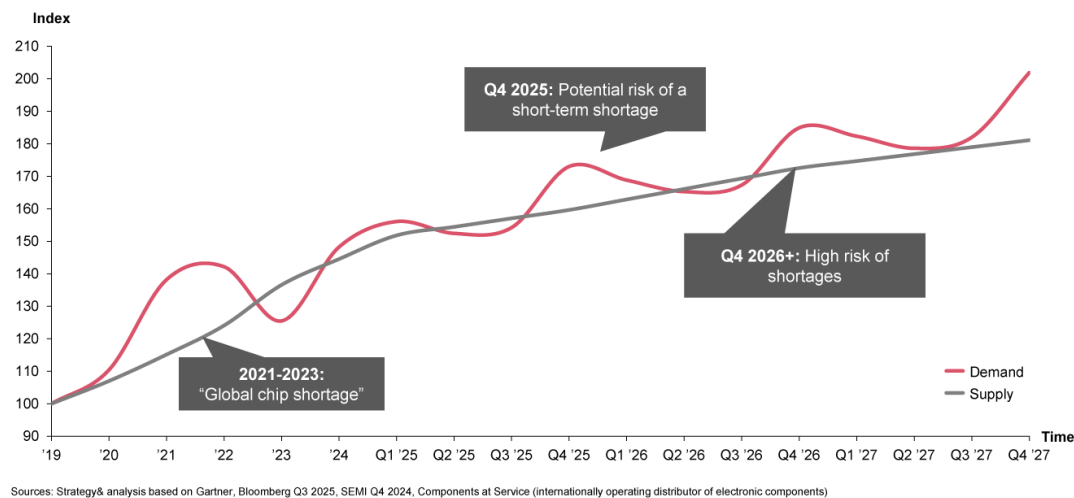


Figure 14.10: Comparison of semiconductor supply and demand

Within the period from 2021 to 2023 (cf. Fig. 14.10), the shortage of semiconductor devices resulted in damages of over 100 billion euros for German automobile manufacturers.

For the development of European industry, not only in the automotive sector, the strategic procurement of semiconductor devices as well as the establishment of in-house manufacturing and design capabilities will be of decisive importance.

14.8 Summary

Semiconductors form the hardware basis of modern systems — from sensors and processors to AI applications — and are the central link between the physical and digital worlds. In particular, the growing demand for MEMS devices (microsensors, actuators) is driving the need for semiconductor manufacturing capacity.

Physically, the function of semiconductors is based on silicon as an ideal carrier material: through targeted doping (n- and p-doping), electrical conductivity can be precisely controlled, giving rise to components such as diodes, transistors, and integrated circuits (ICs). Manufacturing itself takes place through multi-stage processes — lithography for structuring, etching processes for material removal, and ion implantation for electrical activation.

In the Smart Factory, semiconductors are used in sensing, actuation, and control, enabling real-time monitoring, automation, and digital twins. However, the global supply chain (raw material extraction in China, wafer production by companies such as Siltronic, and manufacturing by TSMC/Samsung using ASML equipment) makes the industry vulnerable to disruptions — as demonstrated by the 2021–2023 chip crisis, which cost German automobile manufacturers over 100 billion euros. For European industry, building up in-house manufacturing capacity will therefore be of strategic importance.

CHAPTER 15

Robotics in the Smart Factory

Learning Objectives

After completing this chapter, you will know ...

- ... the significance of robotics in industry and the Smart Factory.
- ... how industrial robots are structured and how they move.
- ... what types of robots exist and what they are used for.
- ... how networked robots collaborate and optimize processes.
- ... how robotics enables flexible and individualized production.

Introduction

Increasing digitalization and automation have shaped industrial production for years and form the foundation for the concept of the Smart Factory. In this context, robotics takes on a central role. Robots are considered a key technology for making production processes more efficient, more flexible, and of higher quality.

15.1 Historical Development and Terminology

Its development has a long history: the first industrial robots were already introduced in the 1960s. These were modeled on the human arm and designed as universal machines, with the goal of automating recurring manufacturing processes without human labor.

To this day, the main objectives of deploying robots lie in economic rationalization and ensuring consistently high product quality, independent of human performance. This is particularly evident in automotive body-in-white production, which, as an almost fully automated field, has become the benchmark for many other industries. Besides the automotive industry and its suppliers, industrial robots are also widely used in the electrical, metal, chemical, and food industries. While their use was traditionally most pronounced in high-wage countries, China is today considered the largest robot market worldwide, followed by countries such as South Korea, Japan, the USA, and Germany.

There is no universally valid definition of the term “robot,” which is mainly due to the diversity of robot types and their application environments. ISO 8373:2021 defines an industrial robot as “an automatically controlled, reprogrammable, multipurpose manipulator, programmable in three

or more axes, which can be either fixed in place or mobile for use in automation applications in an industrial environment.”

Robots are increasingly becoming a fixed part of everyday life and contribute significantly to the further development of intelligent, networked production systems. In the Smart Factory, they enable close cooperation between humans, machines, and digital systems — a decisive step toward Industry 4.0.

15.2 Fundamentals of Robotics

The structure and operation of an industrial robot can be illustrated using the six-axis articulated robot arm as an example, since it represents the most widely used type in industry. In its construction, the articulated robot arm is oriented on the human musculoskeletal system. Its structure is modeled on the human arm. The rotating main axes give the robot its high mobility and enable free positioning and orientation in three-dimensional space.

With a total of six degrees of freedom, the articulated robot arm can reach almost any point within its working envelope. This enables it to execute complex motion sequences precisely and repeatably. Its movements comprise translations and rotations in all spatial directions, which makes it particularly versatile. This flexibility is a key reason for its widespread use in industrial practice.

An articulated robot arm consists of several basic components that together ensure its mobility and functionality. These include the robot arm, the central hand (wrist), as well as the forearm link and the carousel, which are responsible for the system’s movements in space.

The base frame forms the stable foundation of the robot, while the counterbalance offsets the loads during movement and supports precision. The robot controller coordinates the axes and motion sequences, controlled via the teach pendant. Finally, the mounting flange connects the end effector or additional tools to the central hand.

Together, these components form the functional unit of an articulated robot arm and enable the precise execution of complex movements in industrial applications.

The movements of an industrial robot are realized through main and secondary axes. The main axes (A1–A3), also called “major axes,” serve to position the end effector and determine the working envelope. Their movements are linear or rotational and enable the robot’s macro movements. Secondary axes (A4–A6), also called “minor axes” or “wrist axes,” handle fine alignment and orientation in the gripping area. They are almost always rotational and enable precise fine positioning.

Axis numbering begins at the base of the robot. Axis 1 is the first axis of motion at the mounting surface. From there, numbering continues along the kinematic chain to the end effector, i.e., the mounting flange. This numbering also applies to overhead-mounted robots.

The degree of freedom (f) describes the independent movement possibilities of a body. In three-dimensional space, a rigid body has six degrees of freedom: three translations along the x-, y-, and z-axes, as well as three rotations about these axes. Industrial robots generally have six axes, enabling them to reach any position and orientation in space. A higher number of axes allows for greater freedom of movement and more flexible path control, but requires greater mechanical and electrical effort in construction and control.

Robots with fewer than six degrees of freedom are considered globally degenerate. They have a restricted range of motion and allow only limited orientation of the tool. Standard industrial robots, on the other hand, have six degrees of freedom, which gives them full freedom of movement in three-dimensional space and allows precise positioning and orientation of tools. Robots with more than six degrees of freedom are referred to as redundant or overdetermined. They offer more movement possibilities than would be necessary for the task alone, allowing them to flexibly avoid obstacles and execute complex motion sequences.

15.3 Applications of Robots in Industry

Industrial robots have established themselves as indispensable assistants in industrial production, especially where precise, fast, and repeatable work steps are required. Robots are capable of moving loads weighing several tons and performing tasks that are unpleasant, dangerous, or monotonous for humans. Today, industrial robots are mainly used in fully automated mass production. A key area of application is the automotive industry. In body-in-white manufacturing, robots handle almost all repetitive and physically demanding tasks, while in final assembly humans continue to perform the majority of the work, with robots used in a supporting role. Besides the automotive industry, industrial robots are also widely used in consumer goods production and electronics manufacturing.

Further tasks include welding, lifting, transporting, painting, palletizing, machine loading, deburring, and gluing. There, they take on assembly-line work, physically demanding tasks, as well as measuring and inspection applications.

With tasks such as gluing as well as measuring and inspection, they make a decisive contribution to increasing productivity, safety, and quality in modern manufacturing processes. A current trend shows the growing use of assembly and inspection robots, further increasing the efficiency and precision of production processes. In certain areas of automotive manufacturing, the degree of automation already exceeds 95 percent.

15.4 Types of Robots in Industry

Industrial robots differ not only in their task but above all in their kinematics, design, and degrees of freedom. Depending on the area of application, different robot types are used, each with specific strengths and limitations. The following section explains in more detail which types of robots are most significant in industrial production, how they differ from one another in design, and for which tasks each is best suited.

Gantry robots belong to the kinematically underdetermined systems and have three to four degrees of freedom. Their movements occur along linear axes, allowing them to precisely cover large working envelopes while keeping the floor space free. Due to their high load capacities and precise positioning, they are particularly suitable for machine loading, multi-machine operation, the assembly of large assemblies such as engine manufacturing, and the linking of material flows in production lines. Variants such as linear, cantilever, or area gantries allow flexible adaptation to different production requirements.

Palletizing robots specialize in stacking pallets, boxes, or packages and typically have four degrees of freedom. Thanks to their parallelogram kinematics, the tool always remains parallel to the floor, enabling precise handling of heavy loads at high cycle speeds. Typical applications include the automatic loading of pallets, for instance in the beverage or packaging industry.



Figure 15.1: Palletizing robot

SCARA robots (Selective Compliance Assembly Robot Arm) are characterized by their compact design, small footprint, and fast movements. They have four degrees of freedom — three horizontal rotations and one vertical stroke — and can be installed on the floor, on walls, or on the ceiling. Their high repeat accuracy and short cycle times make them particularly suitable for small-parts assembly, pick-and-place tasks, printed circuit board assembly, as well as sorting and packaging processes.



Figure 15.2: SCARA robot

Delta robots are based on parallel kinematics, in which several jointed arms connect the base to a moving platform. They have three to four degrees of freedom and are very lightweight, fast, and precise. Because their main drives are stationary, they can perform highly dynamic movements, but are suitable only for low payloads and small working envelopes. Typical fields of application are high-speed pick-and-place, packaging, sorting, and assembly, particularly in the food and pharmaceutical industries. In combination with camera systems, delta robots can flexibly detect and handle objects.

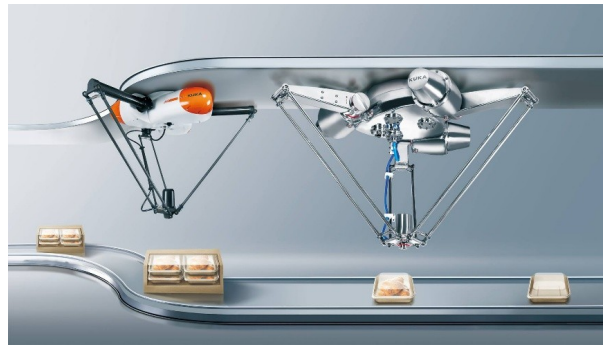


Figure 15.3: Delta robot

Mobile robot platforms differ fundamentally from permanently installed systems in that they can change their position autonomously. They are mainly used in intralogistics, warehouse logistics, and material transport. Mobile robots offer high flexibility, support workers directly at production lines, and do not require a fixed workstation. Modern developments increasingly lead to collaborative systems (“cobots”) that operate without safety fencing and interact directly with humans, making them true cooperation partners in manufacturing.

Each of these robot types is tailored to specific tasks, giving companies the opportunity to specifically optimize their production processes and significantly increase efficiency, precision, and safety in manufacturing.



Figure 15.4: Mobile robot platforms

15.5 Humanoid Robots

Humanoid robots are machines modeled on humans in form, movement, and, to some extent, behavior. They are intended to perform tasks normally carried out by humans, particularly in areas that are dangerous, monotonous, or physically demanding. A current example is Tesla Optimus, also known as Tesla Bot, which is being developed by Tesla to support industry and everyday life.

Tesla Optimus is designed as a fully humanoid robot. Its structure is modeled on the human body to ensure flexibility and versatility. It is about 173 cm tall, weighs between 55 and 60 kg, and has two arms with multiple degrees of freedom, two legs for walking and balancing, as well as a head with cameras, microphones, and sensors for environmental perception. The hands are designed to grip and manipulate objects precisely, while the housing is made of lightweight materials such as aluminum and composites to ensure mobility and energy efficiency. Its power supply is provided by an internal battery, although the exact capacity is not publicly disclosed.

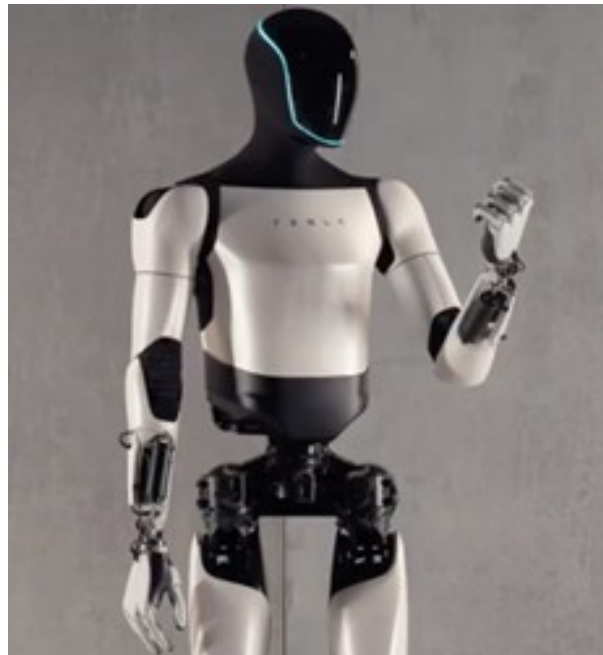


Figure 15.5: Tesla Optimus

The functioning of Optimus is based on a combination of hardware, software, and artificial intelligence. Cameras and sensors capture the environment in real time, while the robot's "brain" is based on Tesla's Full Self-Driving chip and neural networks. This chip was originally developed for autonomous driving and is extremely powerful. This enables Optimus to recognize objects, plan movements, and make decisions autonomously. The 40 electric actuators control the arms, legs, and hands, enabling fine motor movements. Sensory feedback ensures that position, force, and balance are continuously adjusted, allowing the robot to execute movements precisely.

Technically, Optimus offers a range of functions that qualify it for industrial tasks. It can walk at up to 8 km/h and carry loads of 9 kg up to 68 kg, depending on body and arm posture.



Figure 15.6: Tesla Optimus performing a task

In industry, humanoid robots like Optimus could be deployed in various areas. These include assembly and manufacturing, where they could install smaller components, tighten screws, or handle parts. In logistics, they could transport materials, store and retrieve goods, or move products. They could also be used in quality control to inspect products for damage or deviations, and take over tasks in hazardous environments that would be risky for humans, for example in hot, toxic, or radioactive areas.

Mercedes is also among the pioneers in the use of humanoid robots in production. The company is currently trialing the robot "Apollo" at a plant in Hungary, among other things in the areas of assembly, intralogistics, and direct interaction with human employees.

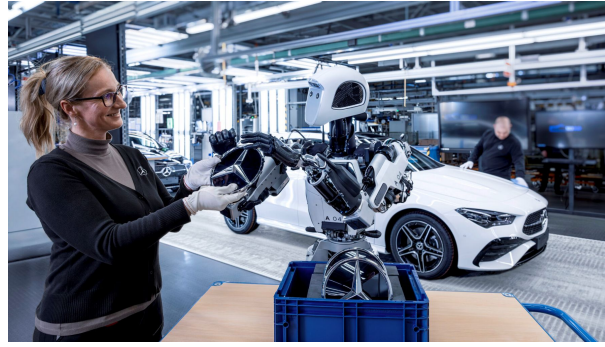


Figure 15.7: Apollo by Mercedes with a human

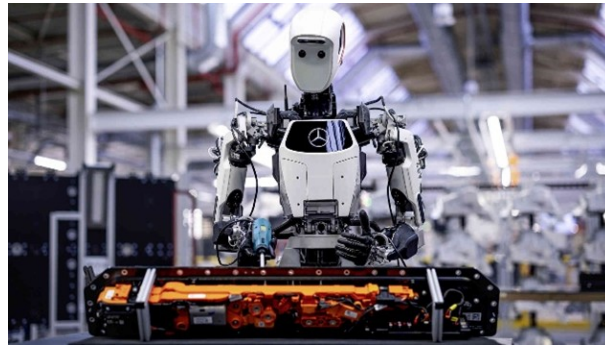


Figure 15.8: Apollo by Mercedes performing an assembly task

The advantages of such humanoid robots lie in their flexibility, the ability to relieve humans of dangerous or monotonous tasks, and easy integration into existing processes thanks to their human-like structure. However, disadvantages include high costs, technical complexity, and autonomy in complex assembly tasks that has not yet been fully achieved.

Overall, Tesla and Mercedes demonstrate how humanoid robots can shift the boundary between human flexibility and robot-based precision. The combination of artificial intelligence, sensory perception, and humanoid design makes it particularly interesting for industrial applications. Even though the technology is still in its infancy and full autonomy in complex manufacturing processes has not yet been achieved, Optimus illustrates the potential of humanoid robots to complement and facilitate human work, and to be deployed in many areas of industry in the long term.

15.6 Internet of Robotic Things

The Internet of Robotic Things (IoRT) extends the concept of the Internet of Things (IoT) to include robots that not only exchange data but can also actively act within their environment. While the IoT networks physical devices such as sensors, machines, or autonomous vehicles and allows them to exchange information, IoRT systems integrate robots that use this data to control their actions autonomously and adaptively. The goal is to form networked robot teams that perform tasks more efficiently, learn from one another, and detect errors. In doing so, IoRT systems connect smart spaces — such as smart rooms, smart factories, smart buildings, or smart cities — with autonomous robots. While smart spaces can monitor states and execute simple actions, robots take on complex activities, such as moving objects or interacting with humans.

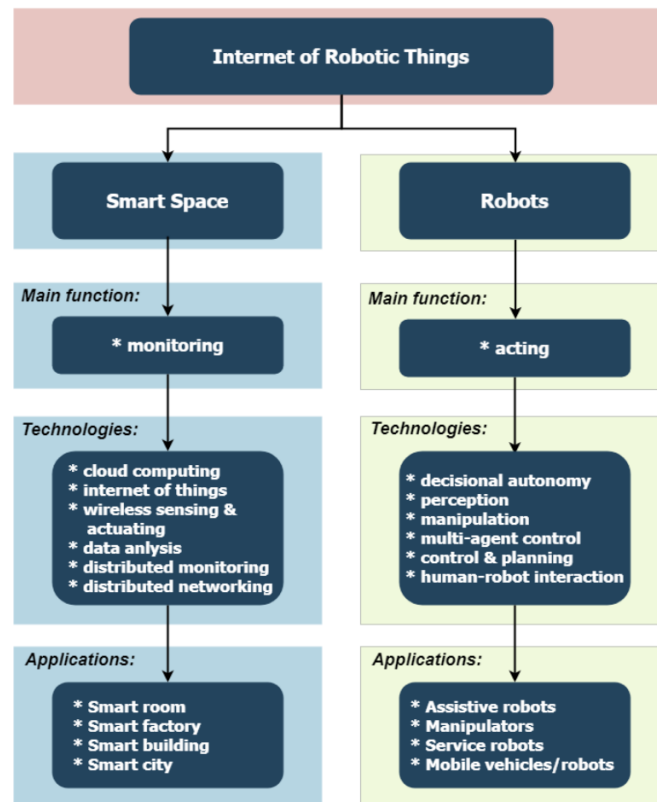


Figure 15.9: IoRT schematic

The advantages of IoRT lie particularly in the fact that robots do not have to develop every cognitive capability themselves but can draw on the information available in the IoT. This allows processes to be optimized locally and in a distributed manner, and classical “machine-to-machine” processes evolve further into “robot-to-robot” interactions. Robots in an IoRT network can share data, learn from mistakes, and identify optimization potential, which leads to efficiency gains especially in production facilities with collaborating robots. Practical examples, such as the Amazon Fulfillment Center, illustrate the benefits.



Figure 15.10: Multi-robot system at Amazon

Here, robots autonomously move storage shelves to employees, detect obstacles via cameras and infrared sensors, navigate using QR codes on the floor, and are coordinated via a central Wi-Fi connection. The result is a significant reduction in long walking distances, faster order processing, and flexible handling of different product types and sizes.

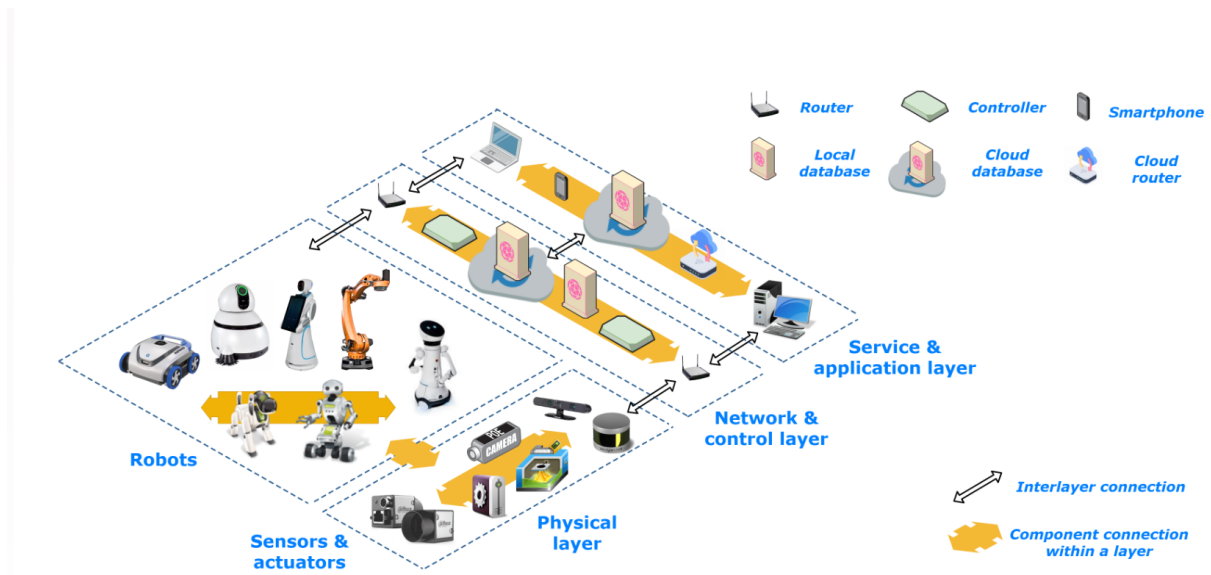


Figure 15.11: IoRT system architecture

The architecture of IoRT can be divided into three layers. In the physical layer, robots act as intelligent agents, communicate with one another, and form multi-robot systems, supported by cameras or motion sensors as well as actuators such as machine tools or 3D printers. The network and control layer comprises routers, controllers, local and cloud databases, as well as the communication and control protocols that ensure coordination between robots, sensors, and actuators. In the service and application layer, programs for monitoring, processing, and control are implemented, which optimize processes with the help of AI and machine learning. Networking and communication often take place via wireless mesh networks to enable data exchange between robots, sensors, actuators, and smart spaces.

IoRT particularly supports multi-robot systems (MRS), in which several robots collaborate within the same environment. The goals here are higher efficiency, distributed task handling, extended functionality, and fault tolerance through redundancy. Challenges lie in coordination, possible interference, cost, and robustness. Multi-robot systems can be classified along various dimensions, such as degree of cooperation, knowledge, type of coordination, team organization, communication mechanisms, team composition, architecture, and team size. IoRT offers a natural solution here for efficiently supporting the network and communication requirements.

Public perception of networked robots is often shaped by fears rooted in science-fiction scenarios in which robots act against humans. In reality, a fully autonomous, human-hostile robot world is pure fiction. Nevertheless, IoRT raises societal questions, for instance regarding potential job losses or the creation of new jobs through AI-supported robot systems. At the same time, IoRT enables flexible cooperation between humans and machines, with robots taking over routine tasks while humans contribute their cognitive abilities, flexibility, and problem-solving skills.

Overall, IoRT represents the next evolutionary stage of robotics. It combines physical robotics with data-based intelligence, enables networked, autonomous, and learning-capable robot teams, and optimizes processes in smart spaces. Practical examples such as the Amazon Fulfillment Center show that IoRT increases efficiency, conserves resources, and lays the foundation for a new, flexible form of production in which robotics and human labor are effectively combined.

15.7 Lot Size 1 in Industry 4.0

Lot size 1 describes the manufacturing of a product as a single unit, i.e., a batch with a size of 1, with the goal of fulfilling individual customer requirements while maintaining economical production. In contrast to traditional manufacturing, where an optimal lot size greater than 1 is sought in order to minimize changeover costs and handle large production volumes efficiently, lot-size-1 production places particular demands on industrial value creation. Due to long setup times and limited flexibility, classic production facilities were only able to manufacture a high variety of variants economically to a limited extent.

The realization of lot size 1 is significantly facilitated by the combination of modern robotics, the Internet of Robotic Things (IoRT), and autonomous transport systems. Industrial robots enable precise, fast, and repeatable work steps that are required for manufacturing individual products as single units. Flexible robot systems allow different production orders to be implemented without long changeover times, while humanoid robots and cobots enable direct cooperation with humans in assembly and during strenuous work steps.

IoRT expands these possibilities by allowing networked robots to exchange information in real time, adapt their actions autonomously, and learn from one another. Integration into smart spaces allows processes to be controlled flexibly, bottlenecks to be avoided, and multiple robots to be coordinated efficiently at the same time. Data from IoRT networks also support the continuous optimization of production and reduce the programming effort for individual machines.

Autonomous transport systems additionally help to make material flows and logistics processes within manufacturing more flexible. Goods, workpieces, or components can be transported between production stations as needed and without human intervention, minimizing waiting times and reducing the throughput time of individual orders.

In combination, these technologies create the prerequisites for economical, flexible, and individualized manufacturing. Automation, networking, and adaptive control make it possible to produce products in lot size 1 without losing efficiency or quality. The Smart Factory thus forms the basis for responding to customer requirements quickly, precisely, and in a resource-efficient manner.

15.8 Summary

Robotics forms the backbone of the Smart Factory and, through flexible, precise, and repeatable work steps, enables the automation of industrial production processes. Different robot types, from articulated robot arms through SCARA and delta robots to mobile platforms and humanoid robots, cover a broad spectrum of tasks, ranging from the assembly of small components and palletizing to direct cooperation with humans. Humanoid robots such as Tesla Optimus or Apollo illustrate how human-like flexibility can be combined with robot-based precision to also take on complex and physically demanding tasks.

The Internet of Robotic Things (IoRT) extends these possibilities by allowing robots to work networked in real time, exchange data, and learn from one another. Integration into smart spaces enables adaptive control, optimizes multi-robot systems, and reduces the programming effort for individual robots. Practical examples such as the Amazon Fulfillment Center show that networked robots can increase efficiency, optimize processes, and improve resource utilization.

These technologies create the basis for implementing lot size 1 in Industry 4.0. The combination of robotics, IoRT, and autonomous transport systems makes economical and flexible manufacturing of individual products as single units possible. The Smart Factory thus enables fast, precise, and

resource-efficient production, in which customer requirements can be integrated directly into the manufacturing process without losing efficiency or quality.

Bibliography

- [1] E. Westkämper. *Digitale Produktion*. Springer Vieweg, Berlin, 2013. (Zitiert auf Seiten)
- [2] T. u.a. Bauernhansl, editor. *Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung, Technologien, Migration*. Springer Vieweg, Wiesbaden, 2014. (Zitiert auf Seiten)
- [3] Birgit Vogel-Heuser and u.a. Bauernhansl, editors. *Handbuch Industrie 4.0: Produktion, Automatisierung und Logistik*. Springer NachschlageWissen. Springer Fachmedien Wiesbaden, Wiesbaden, 2015. (Zitiert auf Seiten)
- [4] H. Kagermann. *Umsetzungsempfehlungen_industrie4_0*. 2013. (Zitiert auf Seiten)
- [5] Litzel Tutanch. *Was ist das internet of things?* 2016. (Zitiert auf Seiten)
- [6] u.a. Vogel-Heuser. *Handbuch Industrie 4.0 Bd.1: Produktion*. VDI Springer Reference. Springer Berlin Heidelberg, Berlin, Heidelberg, 2nd ed. edition, 2016. (Zitiert auf Seiten)
- [7] D. u.a. Lucke. *Ipa_strukturstudie_industrie_4.0_bw*. 2014. (Zitiert auf Seiten)
- [8] Berthold Heinrich, Petra Linke, and Michael Glöckler. *Grundlagen Automatisierung: Sensorik, Regelung, Steuerung*. Springer Vieweg, Wiesbaden, 2015. (Zitiert auf Seiten)
- [9] A. Brandt. *wmr_wirtschaftsbericht2015_230216*. 2016. (Zitiert auf Seiten)
- [10] M. Fratzscher. *Kein roboter kann die Altenpflegerin ersetzen*, 2018. (Zitiert auf Seiten)
- [11] Reiser, Martin, GS1 Germany GmbH. *Epc-informationsservices (epcis)*. 2007. (Zitiert auf Seiten)
- [12] Busse, J. u.a., *Plattform Industrie 4.0. Interaktionsmodell für industrie 4.0-komponenten*. 2014. (Zitiert auf Seiten)
- [13] Geisberger, E., Broy, M. *agendacps: Integrierte forschungsagenda cyber-physical systems. acatech – Deutsche Akademie der Technikwissenschaften. Springer.*, 2012. (Zitiert auf Seiten)
- [14] Gurzki,T., Hinderer, H. *Eine referenzarchitektur für software zur realisierung von unternehmensportalen. Professionelles Wissensmanagement - Erfahrungen und Visionen*, 2003. (Zitiert auf Seiten)
- [15] VDI/VDE. *Statusreport-referenzmodelle-2015-v10*. 2015. (Zitiert auf Seiten)
- [16] Donald A. Norman. *The design of everyday things*. Basic Books, New York New York, revised and expanded edition edition, 2013. (Zitiert auf Seiten)
- [17] Robin Klostermeier, Steffi Haag, and Alexander Benlian. *Geschäftsmodelle digitaler zwillinge*. In Siemens Schweizer AG, editor, *Geschäftsmodelle digitaler Zwillinge*, pages 1–2. Springer Vieweg, Wiesbaden, 2020. (Zitiert auf Seiten)
- [18] Fei Tao and Meng Zhang. *Digital twin driven smart manufacturing*. Academic Press, London, United Kingdom, 2019. (Zitiert auf Seiten)

- [19] Konica Minolta. Digital twins: creating new innovation opportunities, o.J. (Zitiert auf Seiten)
- [20] Siemens Schweizer AG, editor. *Geschäftsmodelle digitaler Zwillinge*. Springer Vieweg, Wiesbaden, 2020. (Zitiert auf Seiten)
- [21] Moritz Schele and Markus Kühn. Vorsprung durch digitale zwillinge: Verknüpfung eines realen produkts mit seinem digitalen abbild. (Zitiert auf Seiten)
- [22] Mindsquare GmbH. Digital twin, 20.8.2019. (Zitiert auf Seiten)
- [23] Raphael Wagner, Benjamin Schleich, Benjamin Haefner, Andreas Kuhnle, Sandro Wartack, and Gisela Lanza. Challenges and potentials of digital twins and industry 4.0 in product design and production for high performance products. *Procedia CIRP*, 84:88–93, 2019. (Zitiert auf Seiten)
- [24] Clarissa Haller. Digitalisierung in der industrie: Zwillinge mit potential, 2018. (Zitiert auf Seiten)
- [25] Sven Altenburg, Hans-Paul Kienzler, and Alex auf der Maur. Einführung von automatisierungsfunktionen in der pkw-flotte: Auswirkungen auf bestand und sicherheit. (Zitiert auf Seiten)
- [26] J-p De Clerck. Insurers want drivers in the vehicle data ownership seat, not car manufacturers. *i-SCOOP*, 1.12.2017. (Zitiert auf Seiten)
- [27] Carlos Miskinis. Practical examples on how digital twin can be used, 2018. (Zitiert auf Seiten)
- [28] Job Wizard. Digital twins: creating new innovation opportunities, 9.3.2020. (Zitiert auf Seiten)
- [29] David W. Cearley, Brian Burke, Samantha Searle, and Mike J. Walker. Top 10 strategic technology trends for 2018. (Zitiert auf Seiten)
- [30] ifaa – Institut für angewandte Arbeitswissenschaft e. V., editor. *Nachhaltigkeitsmanagement - Handbuch für die Unternehmenspraxis: Gestaltung und Umsetzung von Nachhaltigkeit in kleinen und mittleren Betrieben*. ifaa-Edition. Springer Vieweg, Berlin and Heidelberg, 2021. (Zitiert auf Seiten)
- [31] Nachhaltigkeitspolitik: Sieben fragen. sieben antworten., 2016. (Zitiert auf Seiten)
- [32] 10 gründe, warum sich nachhaltigkeit für unternehmen lohnt, 2016. (Zitiert auf Seiten)
- [33] Nachhaltige produktion, 2023. (Zitiert auf Seiten)
- [34] Produktlebenszyklus, 2023. (Zitiert auf Seiten)
- [35] Anne Wiegert. Nachhaltigkeit im supply chain management, 2020. (Zitiert auf Seiten)
- [36] Kreislaufwirtschaft: Definition und vorteile, 2023. (Zitiert auf Seiten)
- [37] Stoffstrommanagement, 2022. (Zitiert auf Seiten)
- [38] Ulrike Böhm, Alexandra Hildebrandt, and Stefanie Kästle, editors. *Klimaneutralität in der Industrie: Aktuelle Entwicklungen, Praxisberichte, Handlungsempfehlungen*. Springer Gabler, Berlin and Heidelberg, 2023. (Zitiert auf Seiten)
- [39] Vogel Communications Group. Mm logistik: Was ist logistik 4.0? alles zum thema digitalisierung und logistik. 2020. (Zitiert auf Seiten)
- [40] F. Heistermann et. al. Digitalisierung in der logistik - antworten auf fragen aus der unternehmenspraxis (positionspapier). *Bundesvereinigung Logistik (BVL) e.V.*, 2017. (Zitiert auf Seiten)

- [41] P. Beisswenger R. Recknagel, editor. *Trotz Control Tower und Digitalisierung - Der Transportplaner bleibt*. Logistik - die unterschätzte Zukunftsindustrie: Strategien und Loesungen entlang der Supply Chain 4.0. Springer Gabler, Wiesbaden, 2., erweiterte und bearbeitete auflage edition, 2017. (Zitiert auf Seiten)