



# **A SemI40 Project Prospective**

**Industry4.0 from Evolution to Revolution**

**SemI40 Consortium**

Copyright © 2019 Semi40 Consortium Partners

PUBLISHED BY SEMI40 CONSORTIUM

[HTTP://WWW.SEMI40.EU](http://www.semi40.eu)

Licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License (the “License”). You may not use this file except in compliance with the License. You may obtain a copy of the License at <http://creativecommons.org/licenses/by-nc/3.0>. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

*First printing, July 2019*



## Contents

<b>1</b>	<b>Preface .....</b>	<b>5</b>
<b>2</b>	<b>Articles .....</b>	<b>7</b>
2.1	Continuous Compliance in Cyber Physical Systems	7
2.2	Advanced Hardware-Based Security for the Industrial IoT	17
2.3	Cycle time prediction with machine learning in the wafer production	25
2.4	Easy Sensor Integration in Industrial Environments	30
2.5	Automated Framework Based on Supply Chain Events	36
2.6	MES Challenges	42
2.7	Machine Learning Algorithms in the Semiconductor Industry	46
2.8	Machine Learning and Automated Decision Making	58
2.9	Key Elements of Productivity Enhancement	76
2.10	Identification of Chiller Models for HVAC Management	81
2.11	A Scalable Wafer Dispatching Strategy Based on DP	86
2.12	Impact of Test Wafer Flows in Transportation Systems	91
2.13	Video-Supported Real-Time Analysis of Products and Processes	94
2.14	Generic On-Line Video Anomaly Detection for CMP Wafer Cleaning	98
2.15	Automated-Generated Long Term Simulation Model	106
2.16	Web-Based Virtualization Toolbox for Integrated Visualization of Data	113
2.17	Tasks and Results of Work Package "Swarm Intelligence"	119
2.18	Closing the Gap Between Exploding Amounts of Data	126
2.19	Capturing Industry 4.0 Use Cases for Interdisciplinary Implementation	137

2.20	Exploiting Advances in Video Technology for Use in Smart Manufacturing	145
<b>3</b>	<b>Roadmaps .....</b>	<b>149</b>
3.1	Road Maps - Cyber Physical Production Systems in Semiconductor Eco-System	151
3.2	Road Maps - Agile Facility	155
3.3	Road Maps - Machine Learning and Automated Decision Making	158
3.4	Road Maps - Digitalization and VirtualizationSemi40	160
<b>4</b>	<b>Semi40 Research roadmap .....</b>	<b>163</b>
4.1	Research Roadmap from a Semi40 Workshop	163
	<b>Bibliography .....</b>	<b>173</b>





## 1. Preface

The primary objective of this collection of articles is to demonstrate how European industries and academic partners developed methods and technologies for automation and smart production systems. Another important objective is to provide examples of success industrial applications for use in workshops and in academic courses. Engineers, technicians and science students can use the wide variety of implementations as examples to learn how many different applications and challenges are included in a project like SemI40 and how it is possible to deal with topics such as automation, process optimisation, energy efficiency, swarm intelligence, machine learning for decision making support, advanced transportation and logistics, compliance in CPS,...

The broad selection of articles offers engineers, scientists, technicians, and managers numerous examples of best-in-class practices by their peers. They provide results and information for a better understanding of the potential impact of some Industry4.0 aspects. Engineers, managers and scientists with a modest background in Industry4.0 topics and methodologies can understand most of the illustrated topics.

Dr. Cristina De Luca  
*Coordinator: INFINEON Technologies Austria AG*  
July, 2019

**Acknowledgments**

SemI40 Consortium would like to thank the authors of the articles for their contributions and the generously time spent for this. We are grateful to many colleagues with whom we have worked and with whom the activities have been implemented over three years. A particular thanks to Mr. Johann Massoner, INFINEON Technologies Austria AG, for the excellent work and support during the project preparation and the two years 05/2016 - 04/2018 coordination.

*The work has been performed in the project Power Semiconductor and Electronics Manufacturing 4.0 (SemI40), under the grant agreement N° 692466.*

*The project is co-funded by grants from Austria, Germany, Italy, France, Portugal and ECSEL JU.*

*SemI40 Consortium  
July, 2019*



## 2. Articles

### 2.1 Continuous Compliance in Cyber Physical Systems

*Author: Ani Bicaku; University of Applied Sciences Burgenland*

*Author: Markus Tauber; University of Applied Sciences Burgenland*

*Author: Christoph Schmittner; Austrian Institute of Technology*

*( Industry 4.0 is the trend of automation and digitalization of industries and its impacts are reflected in all aspects of our live. In Industry 4.0, independent entities interoperate to allow flexible and customized production by integrating Cyber Physical Systems into manufacturing and the use of Internet of Things in industrial processes. In the past, the inclusion of different stakeholders often included a manual security assessment or an audit. This is not enough or scalable in a more and more flexible and connective environment. To still be able to assure cooperating parties that individual components are secured and hence reliable when interoperating with each other, we investigate continuous standard compliance verification. To do this, we define standard compliance based on given requirements and existing standards, from which measurable indicator points are derived. Those reflect configurations recommended by security, safety or organizational relevant standards and guidelines. This work highlights the importance of standard compliance in Cyber Physical Systems and provides information about our standard compliance verification framework developed as part of Semi4.0 for such systems.*

#### Introduction

The evolving and the digitalization of industrial systems increases the number of security vulnerabilities due to the security breaches in the cyber domain. Manipulating or interrupting such systems could also affect safety, which can have consequences such as environmental damage, injury or loss of life [20], [182]. In the case of an unsatisfactory situation it is important to show that the system fulfilled the required security and safety conditions. Another aspect is that the collection of sensor data and triggering of actuators requires secure and reliable data and components interacting with each other. Currently, security assessments are conducted manually by auditors, which are based

on existing standards and guidelines. In addition, connectivity, number of stakeholders, complexity and stakeholders or components, as typically for an Industry 4.0 environment, may change over time. This makes it necessary to provide an automated security and safety compliance and formally describe it in a structured manner [21], [214]. Standard compliance is the adherence to a given set of security and safety requirements, represented by measurable metrics, on the use and configuration of systems or any other security, safety or legal mechanism. These measurable metrics should be imposed by standardized bodies to make each system, device or application comply with the standards. To assure that the system is operating in a secure and standard compliant manner a monitoring module is needed, which is responsible for gathering all the required measurements. Thus, we have implemented a monitoring and standard compliance verification framework, which has been designed to support different use cases and viewpoints that should be considered and researched in Industry 4.0.

### Central Descriptions of the Concept, Activity and Experiments

The new technologies and requirements of Industry 4.0 create a new demand for standardization, which plays a key role in improving security and safety across different regions and communities. The Cyber Physical Systems (CPS) play the major role to understand the nature of the industrial landscape. In the last years, different standard organizations have been established, mostly initiated from industry, and have published various standards in different fields and topics. Despite the extensive research [98], [11], [85] and a considerable number of widely accepted security, safety, organizational and legal standards, existing approaches are insufficient to meet the requirements imposed by challenges and issues in Industry 4.0. In order to address the aforementioned concerns, we have implemented the “Monitoring and Standard Compliance Verification” (MSCV) framework to continuously verify standard compliance in an automated manner, as shown in Figure 1. The MSCV framework uses an Evidence Gathering Mechanism (EGM) to collect evidences from the target system via monitoring agents (MA) [19]. Those are simple scripts evaluating whether a specific configuration is applied to the system which would fulfil e.g. a security control in a security standard, or not (i.e. providing binary information).

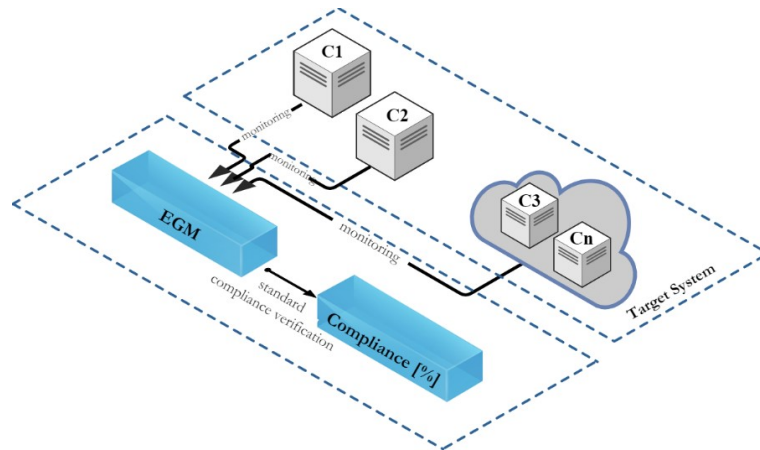


Figure 1: High level view of the Monitoring and Standard Compliance Verification (MSCV) framework

What to monitor and how this links to a standard is formally described in a Measurable Indicator Points (MIP). The MIPs are categorized in Measurable Security Indicators (MSI) and Measurable Safety Indicators (MSFI) and are derived from existing security and safety standards. The EGM sends the necessary information to the compliance module to calculate the standard compliance



including the standard source and a bitwise representation if the metric is fulfilled or not, as shown in the MSCV framework architecture in Figure 2.

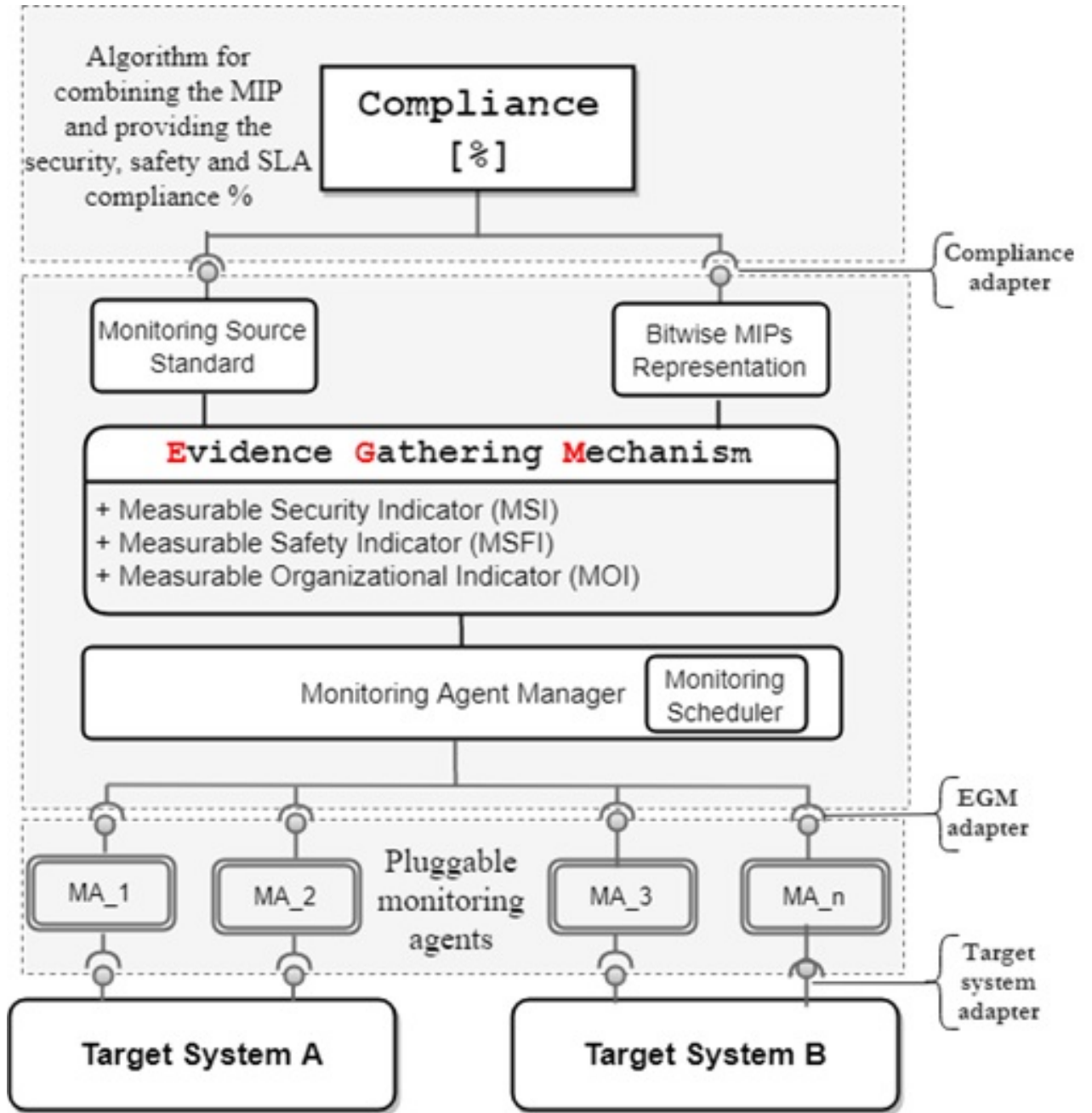


Figure 2: The Monitoring and Standard Compliance Verification Framework (MSCV) Architecture

The standard compliance could be calculated for a single component or for an entire system based on a selected number of metrics from a single/several standard, we however consider aggregation of the information of multiple component to avoid compromising the underlying infrastructure. For each MIP we provide: (i) an ID, (ii) the standard source, (iii) definition (based on the standard), (iv) a monitoring possibility that can be an existing plugin or a customized script, and (v) monitoring value. In order to extract MIPs, we consider the use case depicted in Figure 2, addressing a secure end-to-end communication in CPPS [135]. The data is transmitted between devices, processed throughout the network, and sent to private clouds for further processing and analysis. This use

case will be used as a representative example to better understand the proposed concept and to define and monitor security, business process and safety metrics to enhance trustworthiness in the entire system.

The data is transmitted between devices, processed throughout the network, and sent to private clouds for further processing and analysis. This use case will be used as a representative example to better understand the proposed concept and to define and monitor security, business process and safety metrics to enhance trustworthiness in the entire system. To evaluate the MSCV framework shown in Figure 1, we use the scenario shown in Figure 3 and the OpenStack cloud testbed built in FH-Burgenland. The OpenStack<sup>1</sup> platform makes possible to control large pools of compute, storage and networking resources through a data centre managed from the OpenStack dashboard or via the API. OpenStack works with open-source technologies, which makes it ideal for building, testing and investigating the use case and the MSCV framework.

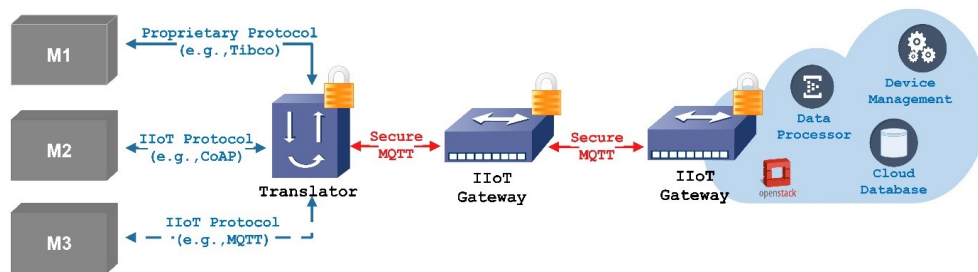


Figure 3: End-to-end communication use case addressing the secure data exchange from the edge devices, such as cyber physical devices (e.g., new and legacy devices in the industrial environment) to the backend infrastructure (e.g., cloud) via IIoT gateways

## Results

The MSCV framework is design to support different use cases and viewpoints that should be considered and researched in Industry 4.0. It allows gathering security and relevant safety and organizational evidence from the target system into a structured way. The architecture of the framework (shown in Figure 2) has a pluggable and expendable architecture allowing easy adaptation to constantly analyse and monitor the status of the system or components of the system. It is able to monitor a large number of MIPs for different CPPS components by aggregating, scheduling, storing, retrieving and analysing the monitoring data to provide standard compliance verification.

To show the standard compliance verification approach, we have considered only MSIs, shown in Figure 4. However, the same approach applies also for MSFIs and MOIs. Each MSI extracted from a standard is monitored using monitoring agents in the corresponding component of the target system.

<sup>1</sup><https://www.openstack.org/>

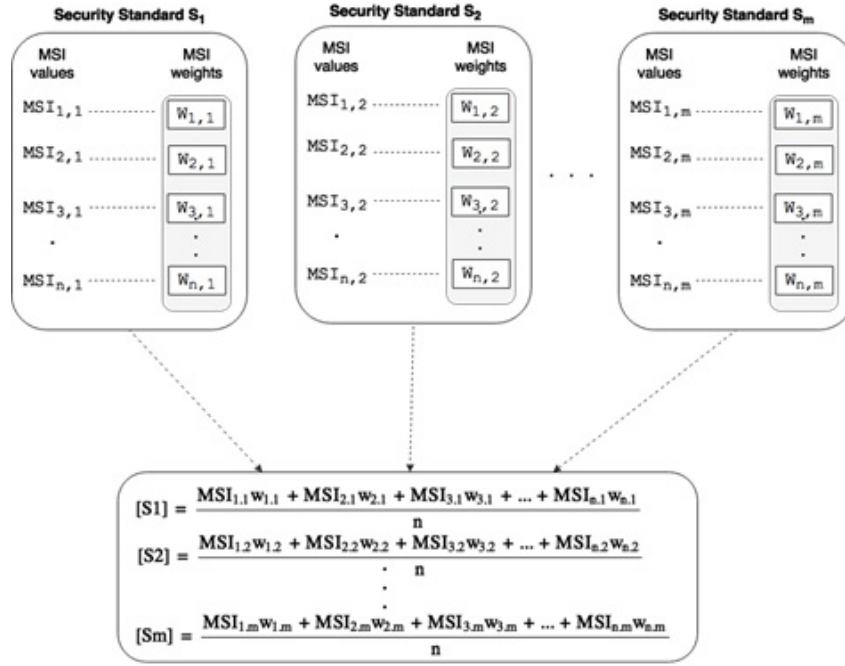


Figure 4: Security standard compliance verification

The monitoring data are then gathered by the EGM module, which is responsible for making them readable for the compliance module. Therefore, the EGM sends the source from which the metric is extracted and a binary value “1” or “0” that indicates if the metric is fulfilled or not. Depending on the specific target system requirements the compliance module assigns to each MSI a weight value to indicate the importance in the range [0, 1].

After gathering all the required evidence from the EGM module, the compliance module first verifies the compliance [%] for a single standard as the ratio between the sum of each MSI measured value multiplied by its weight value and the total number of metrics per standard as shown in Equation 2.1. Then it verifies the total compliance [%] as the ratio between the sum of each standard compliance and the total number of selected standards, as shown in Equation 2.2.

$$MSI_{compliance(j)}[\%] = \frac{\sum_{i=1}^n MSI_{i,j}w_{i,j}}{n} \quad (2.1)$$

$$MSI_{compliance}[\%] = \frac{\sum_{j=1}^m compliance(j)}{m} \quad (2.2)$$

where:

$n$  is the total number of metrics per standard

$m$  is the total number of standards

$MSI_{i,j}$  is the measured value of “ $i$ ” security metric from “ $j$ ” standard

$w_{i,j}$  is the weight value of “ $i$ ” security metric from the “ $j$ ” standard

In order to evaluate the standard compliance a list of representative MIPs extracted from the security, safety and organizational standards is provided. The MIPs are considered in an Industry 4.0 application scenario with the goal to address the requirements of access control systems for the production line. Following are listed three evaluated standards (one for each classification) and a list of metrics addressing the access control requirements in each standard to show how the MIPs should look like in order to be compliant with the MSCV framework.

### MSI - Measurable Security Indicators

#### MSI-1.1: *Secure Identification and Authentication*

- **Source:** IEC 62443-3-3
- **Definition:** The client and the server identify each other and assure their identities via secure log-on
- **Monitoring Plugin:** Can be monitored with Nagios monitoring agent, which checks the configuration of the used protocol (or indeed any other client/server authentication method) to make sure that it uses a secure communication protocol.
- **Monitoring Value:** True/False

#### MSI-2.1: *Strength of Password-Based Authentication*

- **Source:** IEC 62443-3-3
- **Definition:** The system shall be configurable by providing a degree of complexity such as minimum length, variety of characters and password rotation.
- **Monitoring Plugin:** Can be implemented by performing checks on the PAM (Pluggable Authentication Module) to verify if a minimum length or complexity of passwords and password rotation is enabled.
- **Monitoring Value:** True/False

#### MSI-3.1: *Concurrent Session Control*

- **Source:** IEC 62443-3-3
- **Definition:** The system shall restrict the maximum number of concurrent sessions per system account or system type.
- **Monitoring Plugin:** A script can be developed which checks sshd config or pam limits configuration.
- **Monitoring Value:** True/False

### MSFI - Measurable Safety Indicators

#### MSFI-1.2: *Time-Triggered Architecture*

- **Source:** IEC 61508-3, Table A-2, Group 13
- **Definition:** Ensure that the system complies with the safety timing requirements
- **Monitoring Plugin:** Can be checked via Nagios, send test packet to system and check if response time is inside allowed parameters. If Nagios is running on a separate system this achieves medium diagnostic coverage (based on IEC 61508-2). If the systems sends regular information about logical status high diagnostic coverage is achievable.
- **Monitoring Value:** Response time



**MSFI-2.2:** *Techniques and Measures for Error Detection*

- **Source:** IEC 61508-3, Table A-18
- **Definition:** Ensure that system modifications are protected against erroneous
- **Monitoring Plugin:** Check if that system modifications require a password
- **Monitoring Value:** True/False

**MSFI-3.2:** *Control Systematic Operational Failures*

- **Source:** IEC 61508-7
- **Definition:** Ensure that all inputs via a safety-related system are echoed to the operator before being sent to the system. This should also consider abnormal human actions, e.g. speed of interaction
- **Monitoring Plugin:** Can be monitored by a network module that checks the system behaviour
- **Monitoring Value:** True/False

**MOI - Measurable Organizational Indicators****MOI-1.1:** *Event Logging*

- **Source:** ISO/IEC TS 33052
- **Definition:** The system shall forward event log information to a central security information and event management system
- **Monitoring Plugin:** Can be monitored with a Nagios plugin checking syslog/event log configuration
- **Monitoring Value:** True/False

**MOI-2.1:** *Restrictions on Software Installations*

- **Source:** ISO/IEC TS 33052
- **Definition:** The system shall restrict software installation to approved products
- **Monitoring Plugin:** Can be monitored with a custom Nagios plugin checking e.g. paket management (e.g. Linux) or other software management configuration
- **Monitoring Value:** True/False

**MOI-3.1:** *Access to Networks and Network Services*

- **Source:** ISO/IEC TS 33052
- **Definition:** The system configuration must support access to mandatory networks and network services
- **Monitoring Plugin:** Can be monitored with a custom Nagios plugin checking network device and network service configuration (e.g., DNS, DHCP, Gateway, Netmask, NPS, 802.1x Cert etc.)
- **Monitoring Value:** True/False

Each MSI extracted from the standards is monitored using monitoring agents in the corresponding component of the target system. The evaluation scenario takes place in a virtualised environment, where three use case components have been monitored, an industrial device, an IIoT gateway and a cloud database. The EGM collects the data from the target system as shown in Figure 5.

To conclude, for the presented scenario – thus, considering only three main components of the use case, two standards and only a set of representative MSIs, the standard compliance of the target system (end-to-end communication use case) is approximately 75%. Target system components, such as the industrial device and the cloud database need more security controls integrated, whereas the IIoT gateway has already in place most of the required security controls extracted from the standards. Thus, it is possible not only to verify the current standard compliance of the system but also to identify the components, which need more security controls integrated in order to improve the overall compliance of the target system. The same approach applies also for MSFIs and MOIs.

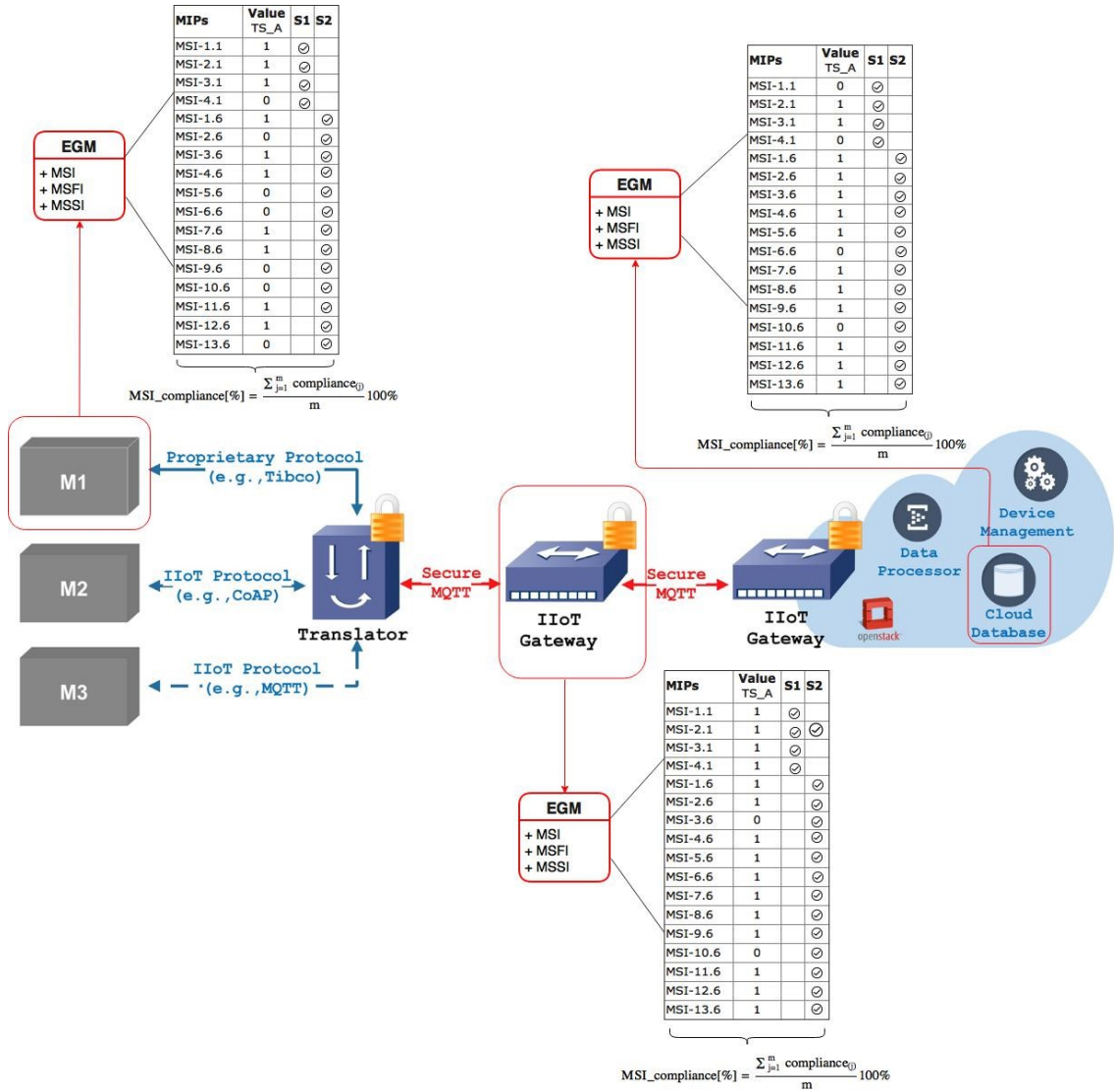


Figure 5: Evaluation Scenario Considering Two Security Standards (ISO 27002 and the IEC 62443)

### Discussion

The MSCV framework presented in this work allows improving security and safety by measuring the standard compliance. The compliance is based on a set of MIPs categorized in MSIs, MSFIs and MOIs extracted from existing standards and best practice guidelines. To show the functionality, the MSCV framework is implemented in an end-to-end communication use case and the security compliance is calculated. As this allows on-demand verification it also affects the runtime engineering and engineering efficiency as compliance checks which had been performed manually in the past, can now be automated. Using the approach throughout the value chain will allow to give information about the degree of standard compliance at any point of time, which may also support legal cases regarding e.g. SLA breaches. Because of the security trust boundaries between for example, different stakeholders, service level agreements (SLA) should be in place.

### Acknowledgements

The work has been performed in the project Power Semiconductor and Electronics Manufacturing 4.0 (SemI40), grant agreement n° 692466. The project is co-funded by grants from Austria, Germany, Italy, France, Portugal and Electronic Component Systems for European Leadership Joint Undertaking.

### About the Authors

**Ani Bicaku**, PhD student at Lulea University of Technology and works as a researcher at the University of Applied Sciences Burgenland in the research field “Cloud and Cyber Physical Systems Security”. Recently, he was working at the Austrian Institute of Technology in the AIT’s ICT-Security Program and was responsible for evaluating data security, data privacy and high assurance in cloud computing. Also, part of his duty was to build an OpenStack Cloud System testbed used for monitoring high assurance of critical infrastructure cloud services. He received the Dipl.-Ing. degree in Communication Engineering from the Carinthia University of Applied Sciences, Klagenfurt - Austria and his B.Sc. degree in Telecommunication Engineering from the Polytechnic University of Tirana, Tirana - Albania.

**Markus Tauber**, FH-Professor at the University of Applied Sciences Burgenland, where he holds the position: director of the MSc program “Cloud Computing Engineering” and leads the research center “Cloud and Cyber-Physical Systems Security”. Between 2012 until 2015 he coordinated the research topic “High Assurance Cloud” at the Austrian Institute of Technology (AIT) part of AIT’s ICT-Security Program. Amongst other activities he was the coordinator of the FP7 Project “Secure Cloud computing for CRITICAL infrastructure IT” and involved in the ARTEMIS Project Arrowhead. From 2004 to 2012 he was working at the University of St Andrews (UK) where he worked as researcher on various topics in the area networks and distributed systems and was awarded a PhD in Computer Science for which he was working on “Autonomic Management in Distributed Storage Systems”.

**Christoph Schmittner**, He received his M.Sc. in System and Software Engineering at the University of Applied Sciences Regensburg in 2013. His main research area is safety and security co-engineering. He works on safety, security analysis and co-analysis methods, connected and safety critical / fault intrusion tolerant system architectures, functional safety and cybersecurity standards and interdependence of safety and security in critical systems. He is member of the Austrian mirror committees for ISO/TC 22 Road vehicles and IEC TC 56 Dependability and designated Austrian expert in corresponding international standardization groups (IEC 61508, IEC 62443 ISO 26262 and ISO/SAE 21434), member of TC65/WG20 “Industrial-process measurement, control

and automation– Framework to bridge the requirements for safety and security”, TC65/AHG2 “Reliability of Automation Devices and Systems” and TC65/AHG3 “Smart Manufacturing Framework and System Architecture” and coordinating the Austrian contribution to the development of ISO/SAE 21434 “road vehicles – cybersecurity engineering”.

## 2.2 Advanced Hardware-Based Security for the Industrial IoT

Author: Sarah Haas; Infineon Technologies Austria AG, Development Center Graz

Author: Rainer Matischek; Infineon Technologies Austria AG, Development Center Graz

Author: Benjamin Bara; Infineon Technologies Austria AG, Development Center Graz

Author: Andreas Wallner; Infineon Technologies Austria AG, Development Center Graz

*The stepwise evolution of today's production facilities into future Smart Factory or Industry 4.0 paradigms entails special requirements on future industrial devices and communication networks. Due to the considerably increasing demand of interconnected devices – for internal and even external remote access – various security measures of future industrial equipment must be improved. An even more challenging use case is the increasing use of industrial mobile robots which will lead to a decrease in work force on the production floor. At the same time, more people on the production floor will be external staff, such as required for robot maintenance, and might be assumed similarly untrustworthy as an adversary. For such advanced use cases, the first part of this article presents specific authentication and authorization concepts, aimed to prevent unauthorized access to mobile robots on the production floor. The second part of this article presents a more practical application and feasibility study of secured communication between today's and future stationary industrial devices – based on the upcoming OPC UA protocol and partially enhanced with hardware-based security.*

### Introduction

One major key factor of typical Industry 4.0 and Smart Factory concepts is the increasing flexibility. This not only includes increased and customizable production flexibility, but also includes requirements for higher flexibility in the way how new production equipment can be installed, regularly maintained, and updated. As pointed out by [86] such flexible Smart Factory approaches enable new business models, services and products, and eventually are expected to increase the economic value of a factory. One prerequisite for the factory to become more flexible is that the interconnections gradually change from the usually centralized approach into a more distributed one. However, particularly due to the increasing demand of extensive machine data acquisition for big data analysis, as well as the increasing demand for device access for remote maintenance (both requiring devices to be permanently connected), various security measures of future industrial equipment and networks must be improved. Consequently, these security aspects are further discussed in this article, with the primary focus on how hardware-based security solutions [80] can be used in addition to improve various security measures of future Industry 4.0 applications. Section II is focused on advanced concepts for future industrial mobile robot applications which are currently not widely established in today's factories – therefore this section is targeted for the long-term Industry 4.0 development. Section III presents a concrete example of an application study how hardware-based security could be already used in a rather short-term Industry 4.0 evolution step for secured communication.

### 1 Industrial Mobile Robots and Challenges

Industrial mobile robots (IMR) are an emerging technology in the industrial context as they provide a high level of flexibility and are able to not only move material but also make decisions autonomously [185]. Mobile robots can move freely in their environment and are easy to replace in the case of a malfunction without the need to stop production in a factory. Despite the many advantages, mobile robots also suffer from drawbacks such as safety issues due to moving alongside humans on the production floor. Safety is an essential problem in the context of robotic and is required to protect humans from any threat coming from mobile robots. For safety reasons, a robot needs to be protected as effectively as possible from manipulations to its soft- or hardware [72].

Security of the mobile robot's components is a prerequisite for safety. Although security is such an important topic to protect robots and humans, research in the area of robotics typically does not focus on security as some more basic topics such as path planning or localization need further research. Due to this fact, the aim of this article is to introduce hardware-supported security measures for mobile robots in the industrial context which also support safety for humans and machines. More specifically, the aim is to prevent access by unauthorized individuals by introducing secured identity checks of robots and people, and to protect the communication between robots and between robots and machines from adversaries. Many different security concepts for securing industrial data already exist but none focus on the field of robotic or use hardware-based security to increase the level of security for their system. Applying existing concepts might provide data security and also the identity of people but those might not be feasible by means of overhead, scalability, real-time-ability, memory usage or performance. Therefore, the first approach of Section II addresses human authentication directly on a mobile robot using one-time passwords (OTP) and is supported by Secure Elements (SE) on the robot and in form of a smart card. The second approach uses sensor fusion and authorization via a machine and a robot to securely authorize action performed by the robot.

## **2 Advanced Authentication and Authorization Concepts for Future Industrial Mobile Robots**

In this section, an offline authentication approach for industrial mobile robots (IMR) using a smart card and biometric data [78], as well as a sensor-fusion based authorization approach for robot actions using Secure Elements (SE) [79] are described. The authentication approach checks that only authorized people gain access to the IMRs. Traditional username and password based authentication approaches are not feasible when accessing a robot on the production floor. To overcome this issue, the propose approach uses a smart card with an integrated fingerprint reader and a Secure Element on the IMR for authentication purposes. The approach is based on time-based one-time-passwords (TOPT). The user needs to place the smart card in the IMR's NFC field and his finger on the fingerprint sensor to start the authentication process. The use of TOPTs is required, so that the values stored on the smart card are valid only for a specific amount of time. The second part is an authorization mechanism which uses the sensor data of both, IMR and machine, to check if the robot is authorized to interact with a specific machine. The combined sensor data is used instead of just the sensor data from the IMR to prevent critical actions from being executed in the case of errors or malicious manipulations on the robot as well as to protect human workforce and production material. If sensor data of the IMR and the machine are valid, the IMR generates an authorization ticket and sends it to the affected actuator. The actuator validates the authorization ticket and executes the command. Each authorization ticket becomes invalid after being used to prevent replay attacks.

### **Authentication**

The authentication approach [78] consists of two steps. In step one the user needs to authenticate himself on the smart card to unlock the computation of the OTP. Step two is to send the computed OTP to the IMR and verify it. In the following passage, the authentication will be described in detail with references to the numbers in Figure 1. (1) Independent from any authentication or other operation, the IMR always computes batches of OTPs, meaning that the IMR computes a number of OTPs for a specific time range  $t$  that can be changed by authorized staff. The IMR might, for example, compute 36 OTPs for a  $t$  of 3 hours where one OTP is computed for every 5 minutes. The computation of batches of OTPs is done to make the verification of an OTP from a smart card possible as the smart card is not able to increase the time due to the fact that it can only compute OTPs when in the range of an NFC field. To compute the OTPs, the time value and the derived key are provided as parameters to a strong cryptographic hash function such as SHA-256. The output

of the hash function represents the OTP. (2) The authentication is initiated by the user who places his finger on the fingerprint sensor and brings the smart card in the NFC field provided by the IMR. (3) The smart card verifies the fingerprint by comparing the input fingerprint with reference data stored on the smart card's memory. (4) The smart card counts the authentication attempts that failed when users provided their fingerprint. If the number of failed authentication attempts is smaller than 10 and the provided fingerprint matches the reference data stored on the smart card, the OTP is computed. The OTP is computed by providing the IMR's derived key and the time value stored on the smart card as parameters for the hash function. The hash is computed by the smart card and then sent to the IMR for verification. (5) In the case that the number of failed attempts exceeds 10 or the provided fingerprint does not match the reference data stored on the smart card, the number of failed authentications is increased by 1 and the authentication is not performed. The notification of a failed authentication is sent to the IMR to inform the user about the failed authentication attempt by the IMR using an LED that turns red. (6) The IMR verifies the OTP by comparing it to the batch-generated OTPs of the last time period  $t$ . (7) The IMR also counts the failed authentication attempts where the provided OTP was not valid. If the OTP from the smart card matches one of the batch-generated OTPs and the number of failed authentication attempts is below 10, the access to the IMR is granted to the user, and the LED on the IMR's NFC module turns green. (8) If the provided OTP by the smart card does not match one of the OTPs from the batch or if the number of failed authentication attempts exceeds 10, the number of failed authentications is increased, and the user is notified about the failed authentication by changing the NFC module's LED to red.

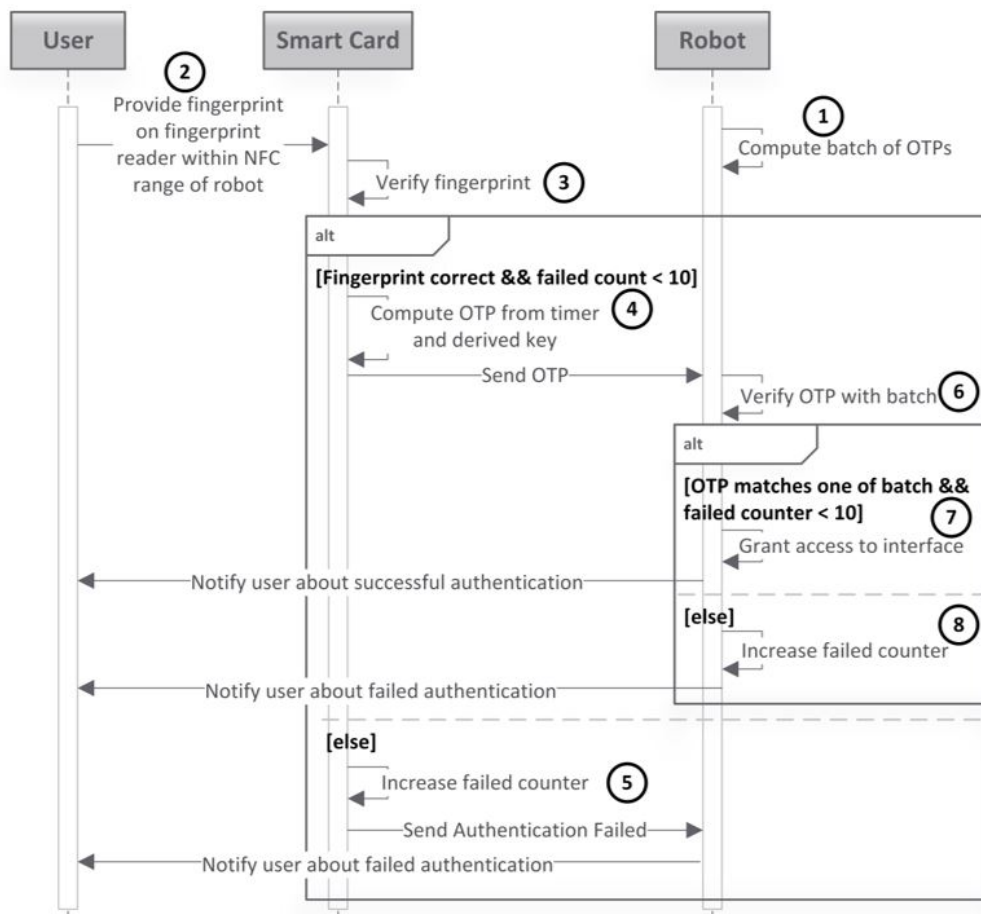


Figure 1: Sequence Diagram of the authentication Approach.

### Authorization

The authorization mechanism [79] uses internal states and the sensor data to check if the IMR is in the correct context and generates an authorization ticket that can be used by the actuator to approve the execution. Figure 2 shows the authorization approach in detail. (1) If the plausibility check returns a positive result or the validity of the incoming command's signature was confirmed, the CCU requests the sensor data from one or more sensors. (2) The sensor(s) compute a signature over their values and return the sensor value and signature to the CCU. (3) The CCU verifies the signatures to check if the sensor values were not modified during transmission. If the signatures are valid, the CCU sends the command to the SE and instructs it to compute an authorization ticket. (4) The SE computes the authorization ticket from the command, secret key, and counter value. The generated ticket is returned to the CCU. (5) The authorization ticket and command are sent to the corresponding actuator's microcontroller for execution. (6) The authorization ticket and command are sent to the actuator's SE for verification. (7) If the ticket is valid for the given command, the SE returns a success message to the microcontroller. (8) In the case of a success message from the SE, the microcontroller executes the requested action.

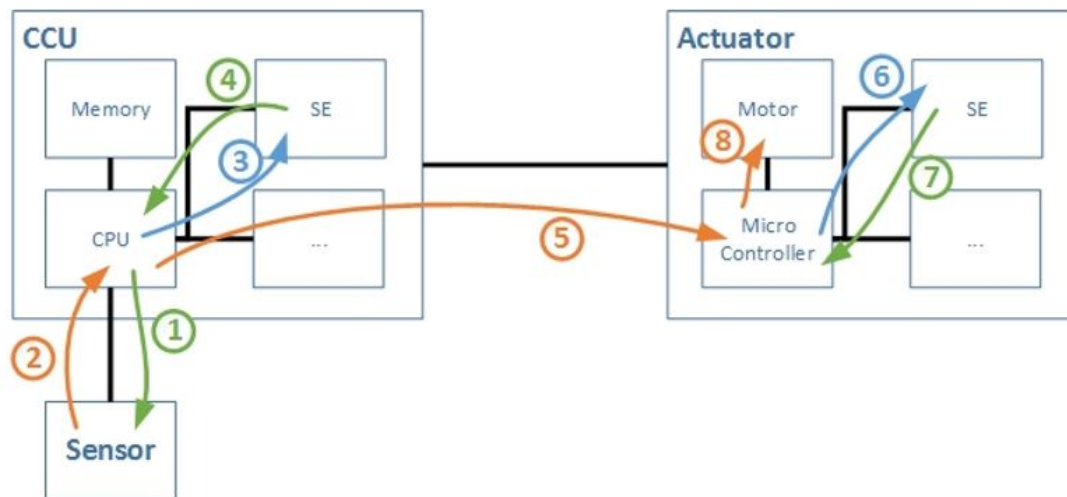


Figure 2: Data Flow when authorizing an Action.

### Evaluation

Both mechanisms were evaluated according to their security features. One of the most common methods to evaluate a security mechanism is a threat analysis [149]. In principle, a threat analysis lists potential threats to a system and shows the countermeasures to overcome these issues. However, not every threat can be prevented and the threat analysis also discovers residual risks that remain even after applying the security mechanism. A threat analysis begins with listing the involved entities such as in our case the IMR, SE and adversary. It also determines the trustworthiness of each respective entity. The next step is to determine the scope of the security analysis by providing assumptions such as the fact that the secret keys are already stored in the SEs. These assumptions are necessary since the security analysis would have to cover much more than the mechanisms directly related to the proposed approach. After defining the scope, the assets that need to be protected are discussed. Since all the necessary definitions and scopes are set, the last step is to list the potential threats and countermeasures that would prevent them. Any threat that cannot be prevented is marked with a residual risk. The severity of the residual risks needs to be assessed by an expert. Depending on the severity and number of residual risks, a mechanism can be considered usable in a given context or not. It again needs to be emphasized that a threat analysis is not



exhaustive and contains only the most important threats to the proposed mechanism. The evaluation of the authentication mechanism (further described in [78]) identified a total of nine threats with no residual risks remaining. Threats include backdoors or wrong/weak cryptography on the SE, loss/theft of smart card, and replay attacks. These and several more threats are prevented by the countermeasures such as expiring TOTP, tamper resistance of the SE or limited authentication attempts. When evaluating the authorization approach, a total of eight threats with one residual risk remaining (further described in [79]). The threats include for example manipulation of commands, manipulation of sensor data, or denial-of-service attacks. The remaining residual risk is related to a denial-of-service attack on the router that connects actuator and CCU of the IMR. This kind of attack cannot be prevented by the given authorization mechanism. Anyways it is not considered to cause a security problem since these attacks simply shut a service down. This means that an authorization cannot take place and the robot would stop working. This issue concerns the productivity but not the security or safety of an IMR.

### **Application Study for Securing Existing Industrial IoT Protocols with Hardware-Security Enhancements**

In contrast to the discussed advanced concepts for future industrial mobile robots – which not yet have been realized in practical applications – this section presents a more practical application, more precisely in the form of an Industrial-IoT-Security application- and feasibility study. The basis of this application study are various concepts published in recent years, proposing the general use of hardware-based security solutions for the Industrial IoT [124], and also proposing the use of these approaches not only to efficiently protect machine identities, but also to protect their communication [80]. However, since most of these recent publications are focused on analyzing the advantages of the security mechanisms itself, current studies lack concrete application examples to document if and how such concepts could be applied to protocols and standards which are either established or upcoming in the industrial context. Therefore, this section presents the concept of a secured Industry 4.0 communication setup and a corresponding implementation based on an industrial demonstration environment which is partly derived from real production environment and protocols. The basic and initial concept of the application study is sketched in Figure 3: Future industrial equipment, such as production machines, robots as well as user-clients for operation staff, is interconnected via existing standard network infrastructure. For security reasons, this concept proposes the use of the upcoming protocol standard for industrial automation “OPC UA” [155], as this protocol also natively supports various state of the art security measures [154]. In addition – in order to further increase the security of the device – the industrial devices are enhanced with a hardware-based discrete Infineon Secure Element. In modern IT- and embedded systems, such hardware-extensions are used as trust anchors for protecting secret keys and cryptographic operations [80], since various incidents in recent years has shown that software measures alone do not generally provide sufficient protection against sophisticated remote cyber-attacks. For illustration purposes, Figure 3 and Figure 4 indicate the additional hardware-based device protection by the green padlock icon. Furthermore, the concept depicted in Figure 3 also considers the step-wise transition of today’s production facilities into Industry 4.0, by integrating existing older “Legacy Equipment” (without built-in security) into the modern secured network with the help of separate “Security Translator” devices. Finally, for enabling future Industry 4.0 use cases – such as remote- and predictive maintenance – another translator device is used as gateway to cloud services (for example via MQTT or MQTTS protocol). In this way, the gateway provides selected filtered machine data, which is only relevant for maintenance purposes, to external companies.

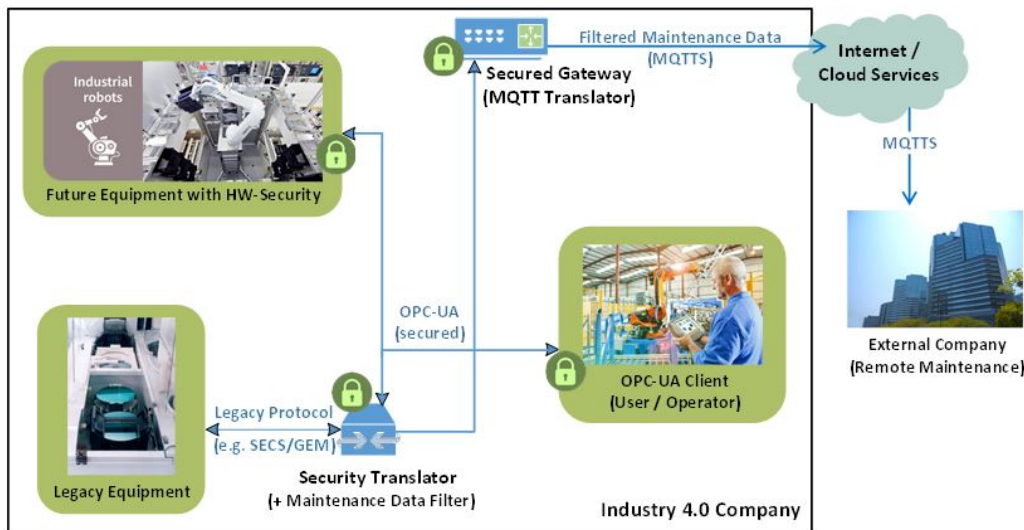


Figure 3: Concept of Hardware-Secured Industry 4.0 Communication Setup.

Finally, in the course of the SemI40 research project, this introduced concept has been implemented and evaluated in a proof-of-concept demonstrator, depicted in Figure 4. Since its primary focus is exploring and evaluating use cases of the discrete hardware-based Secure Element, the actual simulated exemplary industrial data and its communication protocols are based on the common available prototyping components, not intended for productive use (“Raspberry Pi” development platforms and various open-source protocol libraries). The actual focus and novelty of the resulting application study is listed below:

- For evaluation purposes and using higher adaptability, the firmware and API-commands of the Secure Element has been partly enhanced for some derived application-specific use cases.
- For this reason, a customized Infineon test-chip variant (not yet a released product) has been used as basis for this feasibility study. It also includes a specific JavaCard OS variant, and therefore can be partially customized to application-specific needs with JavaCard Applets [96].
- The OPC UA software-library (used as basis on the Raspberry Pi, open62541 [156]) has been partly enhanced to use the Secure Element instead of pure software-libraries, basically for: Generating and storing RSA key pairs and certificates in the secured storage, and checking/signing the certificates as well as session key derivation during the OPC UA session establishment (handshake prior to the data communication).

The demonstration software at the Raspberry Pi devices (and the underlying OPC UA software library) has been implemented in a way to shift parts of the security-relevant keys and operations into the protected execution environment of the customized Secure Element. This includes the initial generation of asymmetric key pairs, secured storage of private keys, and particularly the OPC UA command “Open Secure Channel Request/Response”, where parts of the “Key Derivation Function” (KDF) have been shifted into the Secure Element. Finally, the experimental results were targeted to evaluate the feasibility and also the backwards-compatibility of the hardware-assisted implementation. Therefore, the demonstrator has not only been tested in a pure OPC UA communication setup with partial hardware-based security activated between the OPC UA devices (depicted as green dotted circle in Figure 4), but also tested in a mixed setup, where the partial hardware-protected devices have been communicating with the pure software-based open62541 library implementations. The demonstrated results have confirmed that the hybrid partial hardware-secured implementation is compatible with the original software-based OPC UA

implementation, and therefore can be used for a step-wise integration and migration into future Industry 4.0 methodology, combining:

- Protocol- and security translation of existing older legacy devices (without integrated security)
- First-generation OPC UA compatible devices (pure software-based OPC UA)
- Future potential OPC UA devices with improved hardware-based security measures
- And filtered and secured data translation to cloud services for remote maintenance (e.g. via MQTTS)

All these typical use cases are combined in the application- and feasibility study into an overall future “Industry 4.0 Secured Communication” scenario depicted in Figure 4.

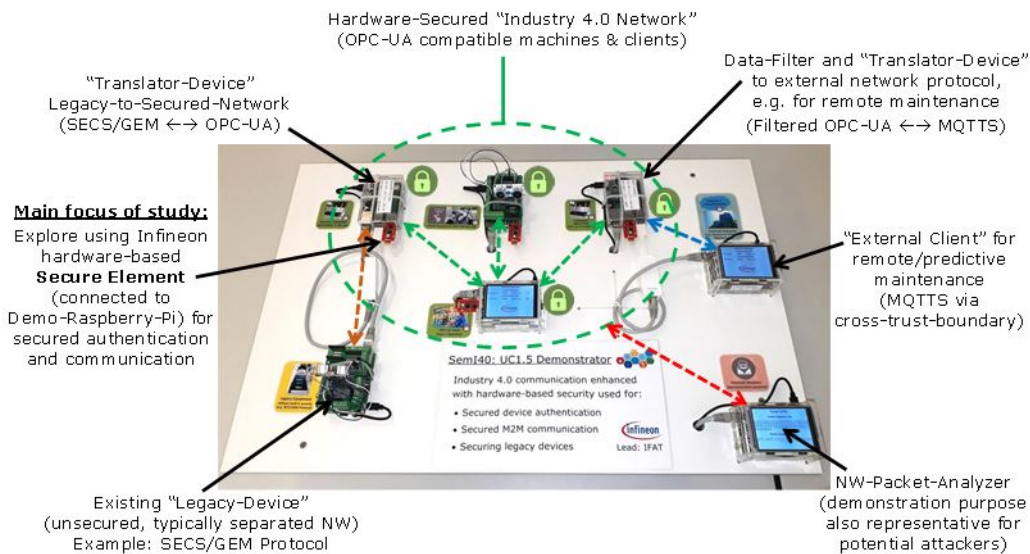


Figure 4: Industry 4.0 Security Application Study and Demonstrator

## Conclusion

This article presented useful approaches how hardware-based security can be used as supplement to existing hardware and software systems, in order to improve several security measures of future Industry 4.0 applications. Depending on the variety of use cases, the proposed hardware-based security approaches and applications can be partially used for authentication and authorization of devices and users, or partially used for industrial communication protocols and protocol translation. The approaches presented in the first part particularly provide safety-supporting security mechanisms required for future industrial mobile robots (IMR). The proposed mechanisms are necessary to prevent unauthorized use of interfaces and to check if an operation by the IMR is allowed. The authentication approach prevents the possibility to change, inject or delete any data such as calibration or configuration data on the IMR. The authorization approach protects human workforce, production material and machines from possible harm due to unexpected movement or interaction. The application study presented in the second part of this article not only provides concepts for integrating and translating existing older legacy devices into modern secured networks. Additionally, the presented implementations demonstrate the feasibility to gradually integrate future hardware-secured equipment in a compatible way into the upcoming OPC UA protocol networks. This approach also allows for the step-wise transition of conventional industrial networks into future secured Industry 4.0 networks.

### About the Authors

**Sarah Haas**, received her Dipl.-Ing. degrees (M.Sc.) in Information and Computer Engineering as well as in Computer Science from Graz University of Technology, both in 2016. The focus in her studies was in Security, Embedded Systems and Sensor Networks, Machine Learning and Big Data Analysis, and Robotics. Since 2016, she has been a Ph.D. student in Information and Computer Engineering at Infineon Technologies Austria AG. Her research interests include security (in particular authentication and authorization mechanisms) of industrial robots.

**Dr. Rainer Matischek**, is a Senior Staff Engineer at Infineon Technologies Austria AG in Graz. He joined Infineon 2004 and first specialized on embedded systems. After his master's thesis in the area of WSN and IoT, in the scope of various EU-funded research projects he focused his PhD research in the area of wireless real-time protocols. Besides a few years working in the firmware product development he finished his PhD in 2011, and finally joined the research department, responsible for various projects in the area of embedded wireless and security systems.

**Benjamin Bara**, is a master student at Infineon Technologies Austria AG in Graz in cooperation with Graz University of Technology. In the course of his master's thesis he focused on applications for hardware-based security and worked on various implementation parts contributing to this article.

**Andreas Wallner**, is a Senior Software Engineer with a main focus on embedded software for security controllers. He studied Microelectronics at the University of Applied Sciences FH Joanneum. His current field of work is in security engineering and operating systems at Infineon Technologies, Austria. His latest publications were focused on authentication for industrial robots and post quantum cryptography on smartcard controllers.

## 2.3 Cycle time prediction with machine learning in the wafer production

Author: Lukas Lingitz; Team Leader, production planning and control

Author: Viola Gallina; Research Assistant, production planning and control

Author: Pedro Vasconcelos; Supply Chain Project Manager, planning and logistics

*Cycle time is one of the most important control parameters of production planning and control. Traditional planning and control methods use static average cycle times. However, real cycle time of a production order is affected by several different factors (such as work in progress on the shop floor or between operations; availability and utilization of equipment, current worker on the workstations or work centers,...) and is dynamic in nature. Only a few approaches target supporting production control decision making based on machine learning predictions. The proposed method uses supervised machine learning algorithms and introduces how the planning quality can be increased with dynamic cycle time prediction based on historical data. The benefits and drawbacks, as well as the application and visualization of the method are also discussed in the present paper.*

### Introduction

In the recent years, artificial intelligence (AI) has become very popular not only in the business and IT sectors, but also in the industry. Machine learning (ML) – as one way of achieving AI in production – is defined as an automated process that extracts pattern from historical data. One of the most frequently applied ML technique is regression that is a supervised ML approach for predicting continuous output variables [107] [116].

Due to the high complexity of integrated circuit manufacturing (such as job shop organized production system with hundreds of different equipment organized in work centers, products with multiple layers and frequent reworks) the semiconductor industry has invested traditionally a lot in advanced IT-systems in the production area, resulting in a large amount, detailed data about the production process. This big data can be analyzed and used for different purposes or predictions for supporting management decision making nearly in real time in today's production systems. Maintenance and quality relating decisions are frequently supported with such analyses and relating tools, however production control applications are not so advanced.

The authors' objective was to investigate the possibilities of cycle time prediction with different regression techniques for supporting production-scheduling decisions.

### Central Description of the Concept, Activity, Experiments

The main objective was to increase the quality of production planning and scheduling based on more accurate target cycle times. The main aim of the authors' activity was to develop a methodology, with which the static cycle times used currently in production planning and scheduling can be substituted with dynamic cycle times depending on different factors.

## Results, Impact, Conclusion and Discussion

### *Random forest*

At the beginning of the analysis, different regression methods were tested on the most important process steps of the wafer production. Our approach is demonstrated on Figure 1. Based on domain knowledge and experience the use case data was statistically analyzed and transferred to the feature engineering and machine learning toolbox. Two different kind of results were achieved at the end of this first stage. On the one hand, a dynamic cycle time prediction with the most accurate model (with the less prediction error) was gained. On the other hand, the most relevant features – filtered from a longer feature list – were determined [158]. After deducing, that based on different aspects (need for data cleaning, training time, accuracy of the model) the most suitable model for this use case could be the random forest model, the frame of the analysis was extended (time frame, analyzed objects and process steps) [128].

### *Dynamic Cycle Time in Production Scheduling*

The cycle time prediction was done for each layer of the 14 most important products. The results are illustrated with the help of data summarized in Table 1 and plotted on Figure 2. In Table 1. the normalized root mean squared error (nrmse) values are given for the currently applied, static planed cycle times (Average) and for the predicted, dynamic cycle times – normalized with the difference between the maximum and minimum observed values. On Figure 2. the real cycle times (black dots) are compared with the currently used, static (orange line) and the predicted, dynamic (green dots) cycle times – for a given product. It must be mentioned, that the machine learning algorithm learns from the past 2000 observations and because of this no prediction is available for these data.

### *Visualisation and Application with R Shiny*

R Shiny – a package from R Studio – is applied for both visualizing and applying the developed cycle time prediction methodology. With the help of this interactive web application, the planner has the possibility to get cycle time prediction for requested lots based on his desired features.

### Tables and Graphs

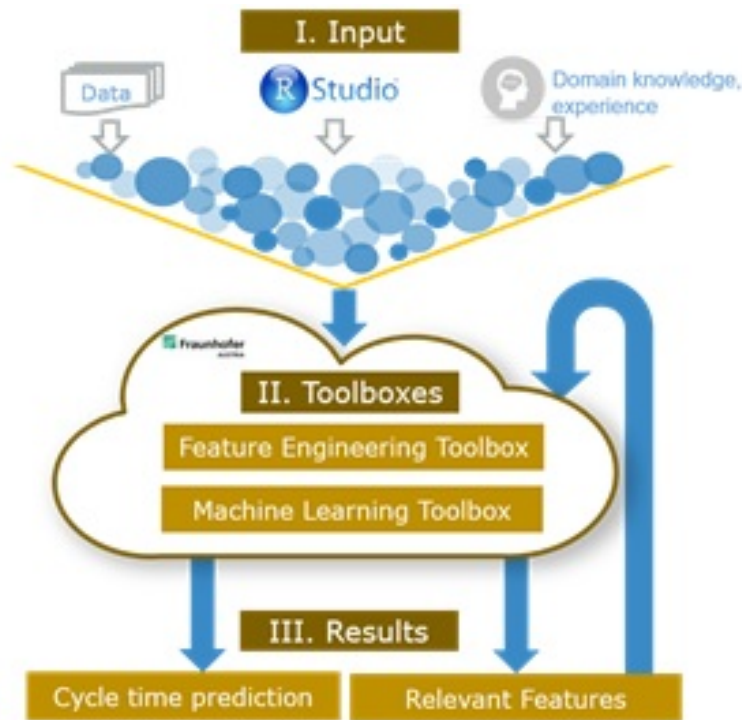


Figure 1: Our approach at the beginning of the analysis

Product	Average	ML	Number of lots
1	17,9	11,7	2228
2	12,5	12,7	1103
3	15,9	13,7	964
4	11,8	18	237
5	23,1	18,4	289
6	20,6	18,8	345
7	23,7	20,4	142
8	27,4	25,1	118
9	25,9	29	32
10	56	41	11
11	37,4	43,4	32
12	19,4	66,3	145
13	18,3	-	7
14	117,8	-	9

Table 1: NRMSE values for the different cycle time calculations



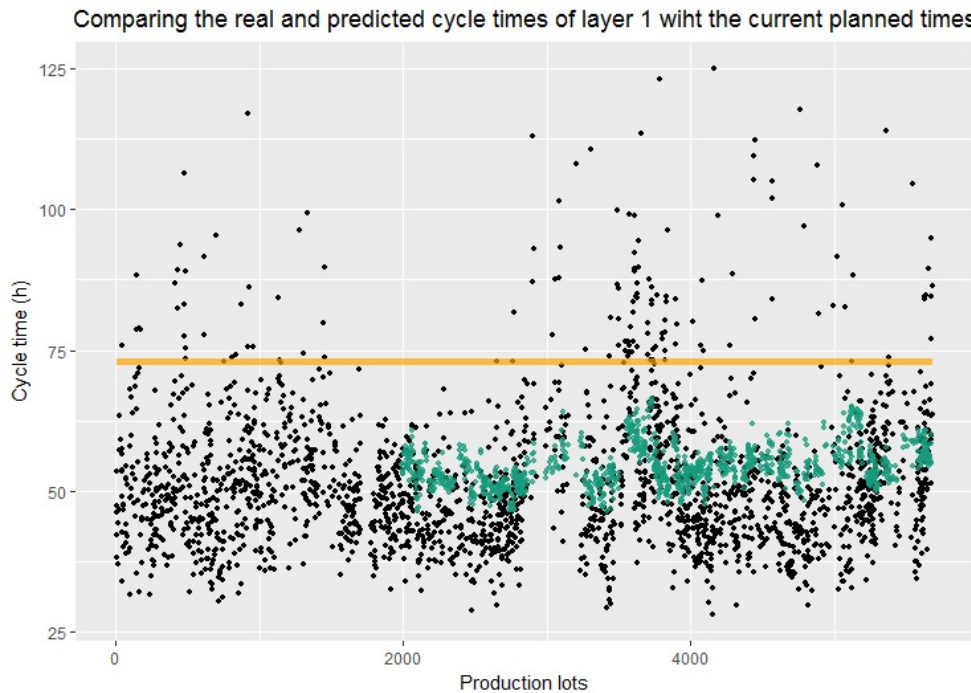


Figure 2: Comparison of the current, static cycle times (orange) with the dynamic ones (green) and with the real cycle times (black)

### Discussion

It can be seen from Table 1, that in the most cases random forest gives a more accurate prediction for cycle time – as our main indicator for the forecast accuracy, the nrmse values are mostly smaller and, moreover, the mean nrmse of the products is 26.5% and 30.5% for dynamic (ML) and static (Average) cycle times, respectively. It must be noted, that the products have various sample sizes and this issue is a sensitive part of a big data analytics project. Because of the few number of lots for the last two products no cycle time prediction with ML could be done. The sample size has a great significance for the accuracy of the prediction too. The more data is available, the accuracy of the prediction could be increased. Including the products with minimum 200 sample sizes modifies the mean nrmse values for 15.5% and 17% respectively.

It must be emphasized, that the presented data is a good example from the 10 various layers of the 14 different products. Almost in all cases ML gives a far better cycle time prediction than the static planned cycle times. This result shows the importance of AI based methods and techniques in the production as well. It has to be mentioned, that a good performing ML algorithm needs a certain amount of data that could be a disadvantage for products with new releases (new products or prototypes) or too few samples. (C products)

### Acknowledgements

The authors wish to acknowledge the financial support of the European Commission for funding the H2020 research project SEMI4.0.



**About Authors**

**Lukas Lingitz** received his master's degree in industrial engineering from the Vienna Technical University. He heads the Department for Production Planning and Controlling at Fraunhofer Austria. His main research fields are production planning, machine learning and Industrial IOT systems.

**Viola Gallina** received her master's degree in engineering management and her Ph.D. degree in production management from Budapest University of Technology and Economics. She supported the SEMI4.0 EU Horizon 2020 Project. Her main research interests include assembly line balancing, linear programming and machine learning.

**Pedro Vasconcelos** received his graduation in industrial engineering and management from the Faculty of Engineering from Porto University. He is Project Manager in the Department for Production Planning and Controlling at Amkor Technology Europe in Portugal. His main work fields are production planning, supply chain integration and project management.

## 2.4 Easy Sensor Integration in Industrial Environments

Author: Yvonne Bergmann; Robert Bosch GmbH

Author: Florian Tripon; znt Richter GmbH

*This document explores the many approaches that can be taken in order to integrate a sensor into an already existing equipment integration environment, together with their limitations and risks. The document also details the main aspects of smart sensor integration into legacy equipment in the scope of Industry 4.0. In this project, we develop an IIoT-Gateway, which is named SemI40 smart Sensor Integration the SemI40 sSI. This paper describes the process of adding external smart sensors to existing elder assets and of the data-transfer to existing or new data repositories. With this use case a solution to improve existing assets management using smart sensors will be shown. We consider two use case applications: (i) the lithography use case with photoresist flow measurement, (ii) the dicing use case with cooling water flow measurement. The main issue when considering data acquisition is which data sources to observe and to which external data sources the acquired data must be linked to. A second issue that must be tackled is how to homogenize the many different data source types so that the acquired data offers additional insight into the actual process details. According to [222] the greatest challenge in Industry 4.0 is standardization. In our project, we have the aim to build the SemI40 sSI that is able to link various external sensors with an equipment's context data. Up until now integrating sensors into existing data landscapes was mainly an on-site, customer specific endeavor. With the rise of Industry 4.0 this will no longer be the case as one of the targets of Industry 4.0 is out-of-the-box process integration. Our research offers an insight into the possibility to bring old equipment with external sensors into the Industry 4.0 landscape. Key to our endeavor was to provide a software abstraction layer, the SemI40 sSI, where sensors measuring different phenomena types (temperature, pressure, flow, etc.) connected on different network types (called fieldbuses) are presented to the user in the same manner – offering the possibility to interact with them as if they are the same. One case study in our project focusses on the dispense detection of a lithography coater tool. The classical approach is to monitor machine parameters such as temperature, pressure or flow at device level. In case that a parameter violates the specification, an alarm will be send out to the tool directly, which still is at device level. The device level is the lowest level according to the pyramid of automation, which is presented by [222]. There is only communication possible between one communication layer to the next-neighboring level, (device-level, control-level, MES-level, ERP-level). In future, there will be a vertical IIoT communication network between these communication levels according to [222]. This is also highlighted by the VDI/VDE status report, which describes the RAMI4.0 (Reference Architecture Model Industry 4.0) that features different hierarchy levels and a value stream [2]. In our lithography case study, we read out dispense values by an external sensor. The SemI40 sSI is needed to link the device parameter resist flow to the context data, i.e. the wafer identification number. This combined data set will be saved on a data server. The combination of the dispense data with a specific wafer number poses a challenge, because our coater tool has no interface like SECS or GEM. An example of a recorded dispense curve featuring five dispenses is shown by Figure1. Looking at the dispense detection from a vertical IIoT approach, we find the following benefits:*

- *Quality control of the dispense spin-coating process*
- *In-situ dispense monitoring saves daily checks, i.e. operator time and machine up-time*
- *If dispense specification limits are violated – direct vertical communication from the **SemI40 sSI** to the MES, which means that material rework can be triggered and the coater tool is stopped by the system, machine checks can be triggered automatically.*
- *In-situ dispense monitoring shows the exact amount of photo resist consumption, hence the consumption data can be communicated to the ERP-system and to the photo resist supplier.*

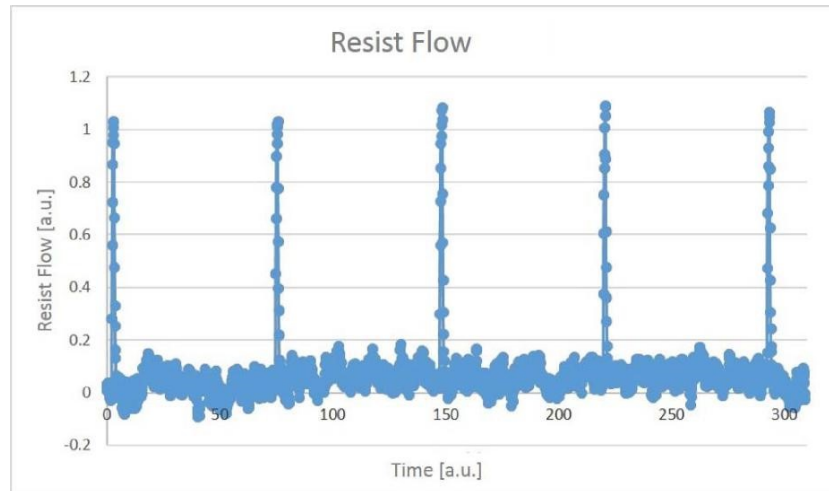


Figure 1: Resist flow measured by the flow sensor vs. time, 5 dispenses

Figure 1 shows the resist flow plotted over time, where five resist dispenses are recorded. The data points have to be combined with the corresponding context data by the IIoT-Gateway. The software demonstrator that was built to prove the findings of the research offers the possibility to create sensor models in our software demonstrator based on sensor description files (mostly offered by the sensor manufacturer). These models are instantiated and added to data collections which are started by external triggers. Recorded data is linked to context data points (which themselves are recorded from other external systems) and sent to data sinks using specific data mappings. Using this approach, in the “Cooling water flow measurement for dicing equipment” use case, we built a model for a Trafag CMP 8270 sensor that uses CAN as communication interface. The recorded data is enriched with context data gathered from Tibco RV messages that the equipment host sends to the MES. The triggers to start and stop the data collection are themselves Tibco RV messages. The recorded and enriched data is mapped and stored in an Oracle Database. Nine sensors are simultaneously linked to the software demonstrator with data being recorded on demand per sensor instance. Figure 2 displays the sensor setup together with a snapshot of the recorded data.

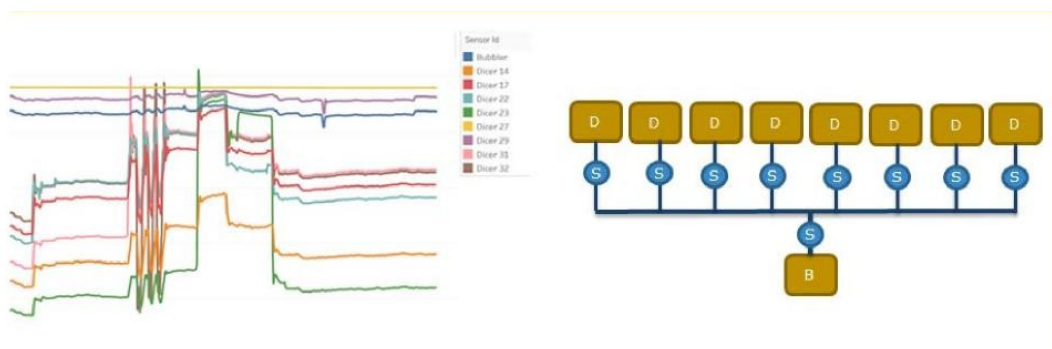


Figure 2: Cooling water flow measurement for dicing equipment

For the “Cooling water flow measurement for dicing equipment” use case, the nine sensors were installed so that pressure data is measured at several dicer tools and from one sensor at a bubbler tool in the main supply tube. Data from the nine sensors is plotted in the left picture. Variations in the water pressure can be observed, which correlate with the start and stop of a processes step at one of the dicer tools.

### Introduction

Smart sensors can deliver various additional information especially if legacy equipment is used in production. That legacy equipment provides only a limited view on the real processes. Thus smart sensors enable manufacturers to collect important information that allows them to improve their production processes beyond state-of-the-art. The challenge of today's IT infrastructure is, that available smart sensors have various interfaces for transferring the information to central systems and huge efforts are needed to integrate such sensors. Especially OPC-UA is getting more popular in the context of Industry 4.0, but no companion standards for the semiconductor industry have been defined similar to the SECS/GEM protocol.

For an easy and quick integration of sensors into an existing IT landscape a highly configurable software system was built that allow users just to configure the software rather than require coding for the integration of sensors. Today's IoT developments have made sensors and communication hardware rather cheap, especially when compared to the cost of the latest semiconductor equipment. It is therefore tempting, to simply add a plurality of sensors for temperature, pressure, gases and the like to existing equipment, to tap the manufacturing product data base(s), and to start "drawing big spreadsheets". Considering the actual number of tools and the number of wafers with different products processed, such a blind, brute-force strategy reaches quickly its limits - even with today's data crunching possibilities - by generating too many events without the additional information necessary for the system that has to effectively decide which action to take.

### Central Descriptions of the Concept, Activity and Experiments

Additional sensors employed at manufacturing equipment can provide substantial improvements in the quality and parameter consistence of the process and the final product, which in semiconductor manufacturing is in most cases a silicon wafer. The main issue preventing the use of additional generic sensors for process control is the absence of an additional sensor interface at the manufacturing equipment, both in HW, but in particular in the SW, which manages the tool. The existence of two different data sources, on one side the equipment with its internal sensor and system management SW connected to the Tool-/APC-Controller, on the other side the sensor with a separate data management system, at least complicates a consistent data analysis, or makes it even impossible, as ad-hoc implemented "solutions" rely heavily on manual and therefore error-prone procedures. In this case, the potential benefit of the use of additional sensors will not materialize.

These most important disadvantages are:

- Relative high effort for system maintenance because of additional data transmitter and additional software
- The overall system is less stable than without a box between equipment and EI
- No possibility of parallel IIoT developments

These supplementary requirements for a sensor interface led to a new sensor integration concept, which is shown in Figure 3 (System Overview).

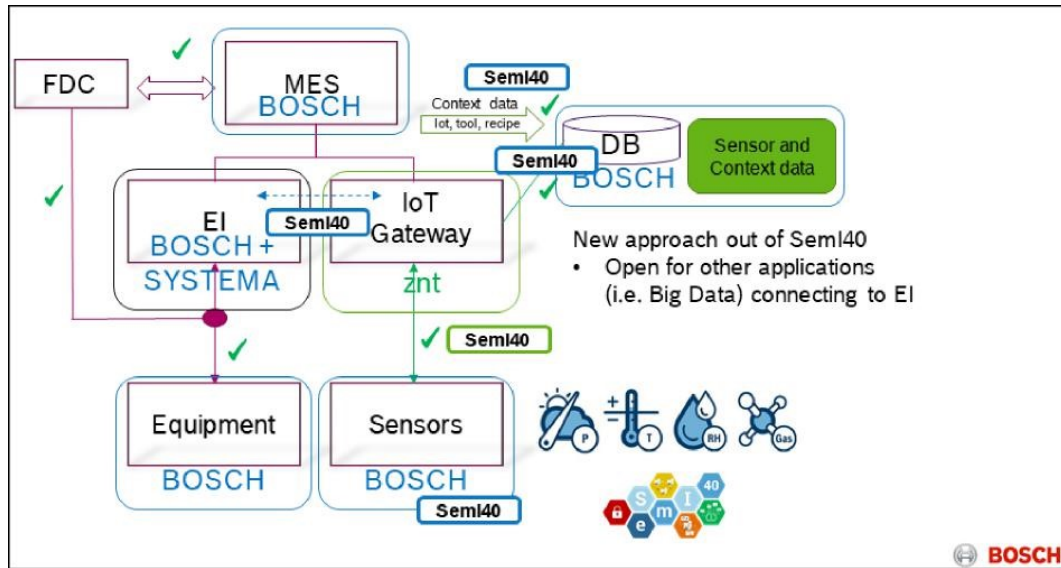


Figure 3: System overview

The new concept for sensor interface has the advantage that the separate Sensor EI will neither disturb, interrupt nor modify the SECS point to point equipment to host connection. Consequently, the stability of the system is ensured and the new concept will not cause higher maintenance effort for the equipment integration path.

In order to realize this new concept, the following additional considerations have to be taken into account:

- **SemI40 sSI** collects context data from the Message Bus
- **SemI40 sSI** assembles the data and associates it to one or more equipment
- One option is to see the **SemI40 sSI** as a process tool with a SECS connection
- A further option is the use of alternative standard protocols such as OPC-UA or MQTT

The **SemI40 sSI** that separates the data sources from the data sinks makes it possible to use simultaneously different sensor types communicating over different networks and different network types.

Figure 4 shows a list of sensors that have already been connected in the scope of this SemI40 WP2 project. Supported sensor types that can easily be modeled in our software demonstrator.



Figure 4: List of sensors that have been connected in the scope of SemI40 WP2

In our test environment we combined on-the-fly simultaneously recorded data from a Trafag CMP 8270 sensor communicating over a CAN interface together with data recorded from a Sonoflow CO 55 sensor communicating over Modbus TCP. Using configured Data Mappings, the recorded data was simultaneously made available in a local SECS stream and also saved in an Oracle Database. In this way we proved that, for the end-user, the source of the data, the hardware used to transport the recorded data and the destination of the data are highly homogenized, completely hiding all the technical differences that make up the myriad of systems that can be linked using the **SemI40 sSI**.

### Results, Impact and Conclusion

As a result we developed the SemI40 smart Sensor Integration (SemI40 sSI), which is a type of IIoT-Gateway. It is a generic platform promoting easy sensor integration for various applications. The impact is, that condition monitoring is enabled for the production tools considered, e.g. the dicing tool or the lithography coater. The benefits are an improved quality control, an increase in tool uptime and a reduced consumption of chemicals.

The software demonstrator proves that it is possible to record data from disparate sources, to aggregate and store them, on demand, in disparate data sinks. The above described process of integrating the sensors can be done by a shop floor technician using only the provided graphical user interface, without having to write a single line of code. It has been shown exemplarily that the desired functionality is guaranteed. Currently, there is no better/other solution known that delivers comparable results.

### Discussion

Monitoring a production tool's parameters poses a challenge on its own. Also, elder equipment show deterioration effects. Sometimes, spare parts are no longer available or a machine type is not supported by the supplier any more. While developing the SemI40 sSI, we found that there were no standards in terms of sensor protocol, e.g. MODBUS TCP or concerning the interfaces between the device, the equipment integration and the MES. For future projects, standards may help to combine the device data with the production materials' context data.

### Acknowledgments

The authors wish to thank all our team members involved in this development. The development was done as part of the JU ECSEL project SemI40 in cooperation of Bosch with znt Richter. The funding is gratefully acknowledged.

### About Authors

**Dr.-Ing. Yvonne Bergmann** received her physics diploma from Berlin Free University in 2010. She joined Robert Bosch GmbH in 2010 and received her Doctor of Engineering in 2015 from Christian-Albrechts-University of Kiel. Since 2013 she has been working in the manufacturing section frontend in the lithography department in the group Equipment Engineering - Stepper/Scanner/Track.

**Florian Tripon, BSc.** graduated in 2006 from the Faculty of Informatics at the West University of Timisoara, Romania. He started working in the semiconductor manufacturing software business in 2007, with experience on building software solution on the Equipment Integrating and on the MES level.

## 2.5 Automated Framework Based on Supply Chain Events

Authors: Dennis Bauer, Andreas Schlereth and Florian Maier; Fraunhofer IPA

Authors: Thomas Ponsignon and Bernd Waschneck; Infineon Technologies AG

*We present in this chapter an automated framework for linking supply chain events to production management. Thereby, the challenge of acting more flexibly in volatile markets for manufacturing companies is addressed. While events in the supply chain are detected using statistical process control (SPC) in the proposed framework, the control loop to production is closed using ideas inspired from run-to-run control. The framework is expected to offer a higher level of flexibility with respect to demand changes. This results in increased on-time delivery for customers. While the semiconductor industry faces the challenge of volatile markets, long cycle times and short product life cycles, systems for demand planning and production management are nowadays decoupled. Therefore, the proposed framework supports semiconductor manufacturing in linking demand planning with production management.*

### Introduction

Volatile markets with short product lifecycles but long development and production times increase complexity and pose challenges for manufacturing companies [15]. The ability to flexibly meet customer demands is becoming a significant competitive advantage. Production planning and control plays a key role to continuously influencing this complex interplay of supply chain and production [229]. To avoid large inventories and possible write-offs, forecasts, production planning and production control need to be closely linked and well aligned. Thereby, production control is responsible for successfully implementing the production plan despite all disturbances [130]. In Industry 4.0, this concept of linking various players in the supply chain is called horizontal integration [22]. It ensures a high transparency, fast reaction times and a cross-system optimization.

The semiconductor industry faces the challenges of long cycle times up to several months, short product life cycles and volatile markets [144]. As early as 2001, the consortium of the International Technology Roadmap for Semiconductors described their vision of Factory Integration to address these challenges [93]. Today, this vision is extended to a global virtual factory, which links the production network with the supply chain planning processes [75].

In this chapter we describe an automated framework for linking supply chain events, e.g., changes in demand planning, to production management. Today planning systems for supply chain and production are often decoupled: after the demand is released to production, it enters into a freeze fence in the planning systems and it cannot be changed. The production commits to a delivery date with respect to the cycle time. During the freeze fence period, demand changes have no impact on production. Therefore, decisions for adaption are unnecessarily delayed and corrective measures in manufacturing may take place too late.

### Concept of the Framework

Our new integrated approach addresses this challenge to increase the flexibility of companies acting in volatile markets. First, supply chain events are automatically detected. Then, these events are incorporated into the dispatching process in complex job shop environments. Therefore, a closed loop between supply chain disruptions and production control is established (cf. Figure 1).



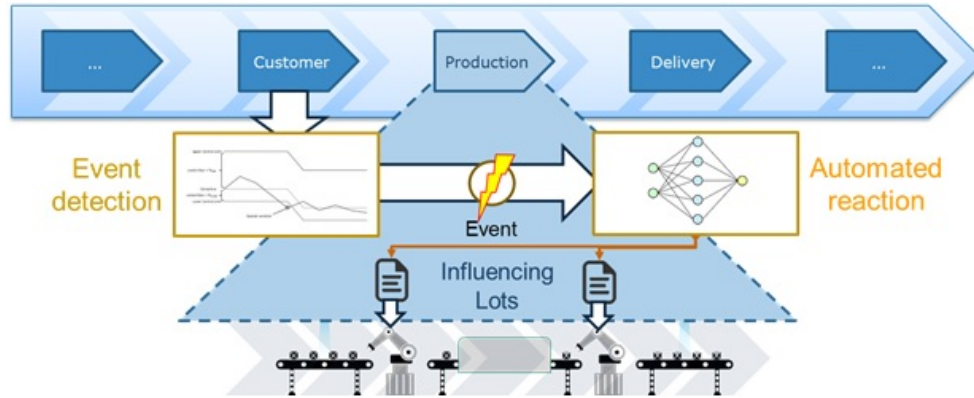


Figure 1: Concept of an automated framework to influence production dispatch based on supply chain events (Bauer et al., 2017)

### *Detection of Supply Chain Events*

The digitalization and globalization of supply chains lead to comparable levels of complexity and innovation speed in the incorporated processes and the manufacturing process alone. In contrast to manufacturing processes, supply chain processes do not necessarily follow a certain, simply assessable, distribution. A major difference is the influence of human interactions in the processes and specific supply chain challenges like the bullwhip effect. Consequently, a simple transfer of SPC from the manufacturing level to supply chain processes is not possible since these specific characteristics have to be considered and integrated into the monitoring concept. Especially the semiconductor manufacturing, with its complex processes and the accompanying challenge of the long cycle times compared to the short life-cycle of the products containing semiconductors, necessitates an accurate monitoring system for all kinds of different supply chain processes.

In semiconductor business, market demand is a crucial source of information for planning supply chain operations. The demand planning process deals with highly volatile customer orders and forecasts. Hence, we intend to apply a SPC approach to monitor demand information and to react upon violations of control limits. A control mechanism shall be developed, which regularly scrutinizes customer demand, and it triggers an appropriate corrective action process in case of a significant deviation. The proposed approach considers the individual ordering behaviour of the customers while generating the control charts. Therefore, the number of false alerts, compared to conventional monitoring solutions, is reduced and a more accurate detection of the relevant deviations is provided in comparison to other today's approaches. Ultimately, the developed concept shall be a lever to reduce the workload of the planners, to smooth the supply chain planning activities, and to improve the product flow through the manufacturing system since relevant deviations can be detected as soon as they occur and the planned quantities can be adjusted respectively to avoid short-term changes associated with high costs. Therefore, we propose a hierarchical framework for applying SPC to demand monitoring. The framework is composed of three levels: Company Level, Customer Level, and Product Level. Each level provides information that empowers the whole SPC approach by using dedicated control charts. Information on a given level can be explained by a related change on the higher level and can be detailed on the lower level. Depending on the purpose of each level, different types of control charts are used, among others X-bar chart, S-chart, moving range chart, and individual control chart. An example of the proposed SPC framework is depicted in Figure 2.

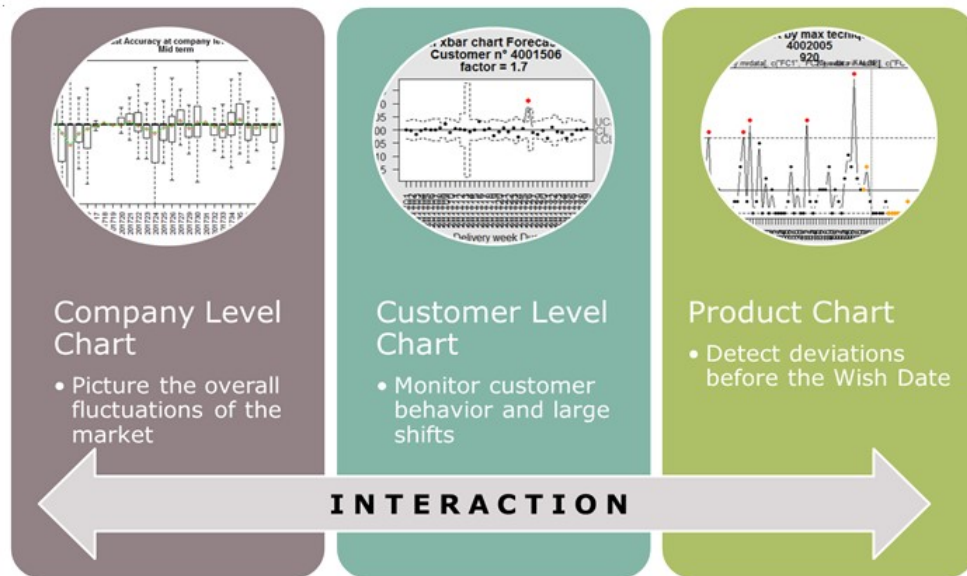


Figure 2: Proposed SPC framework to monitor demand changes in the supply chain

### *Production Dispatch Influence*

Due to criteria such as re-entrant production flows, unequal processing times or sequence-dependent setup times semiconductor manufacturing can be described as a complex job shop. Within complex job shops, scheduling and dispatching highly impacts the performance of the manufacturing process [143]. Therefore, the closed loop is established between supply chain observance (cf. previous section) and dispatching assigning the job with the highest priority from a queue of waiting jobs to the machine [227].

Inspired by the concept of Run-to-Run-Control (R2R) which has been applied to manufacturing processes in semiconductor front-end in the last years with great success [146] we developed a framework in an event-driven system architecture to enable semiconductor manufacturers to react automatically and faster on supply chain events. The so-called Advanced Dispatch Control (ADC) influences production lines based on events observed in encapsulated sections of the supply chain. ADC will connect those different domains of a company to push relevant information from supply chain to production line and determine corrective measures to influence production line autonomously. The approach consists of the following four major steps (cf. Figure 3):

1. Receive an event from a supply chain observer component (e.g. customer request for early delivery).
2. Analyze the received event and retrieve additional context information (e.g. context to identify affected lots) from existing data sources (e.g. Manufacturing Execution System).
3. Decide for corrective measures that can be applied to an existing dispatcher component (e.g. speed up several lots and slow down others) and evaluate them as well as their consequences (e.g. work in progress level, cycle time, etc.) by simulation.
4. Apply correction plan to production by influencing lots via the existing dispatcher component (e.g. Applied Materials' Real Time Dispatcher) and thereby adapt dispatch sequence on manufacturing equipment in production line.

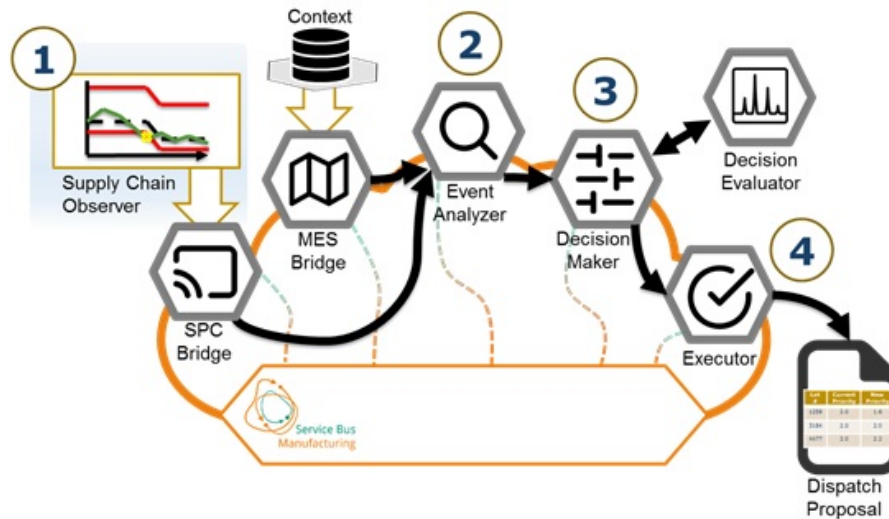


Figure 3: Steps of production influence by Advanced Dispatch Control

Key components of the ADC architecture implementing these steps are:

**Event Analyzer:** The event analyzer receives events from supply chain observer and enriches them with additional data from manufacturing IT systems such as Enterprise Resource Planning or Manufacturing Execution System. While events already need to contain some basic information such as data on the violation of a limit or product relation, manufacturing IT systems need provide additional context information. Among them are data on available priority corridors or current work-in-process of the respective product.

**Decision Maker:** Using the context information supplied by the event analyzer the decision maker determines on one multiple corrective measures for the occurred supply chain event. Therefore, the decision maker correlates with the controller in the R2R concept. While R2R in manufacturing aims at minimizing variability in processes, ADC aims at grading output which is to some extent similar. Decision making is done by using methods of machine learning to identify similar situations in the historical data stock (classification) followed by the determination of the magnitude of influencing parameters (regression). Additionally, the decision evaluator as a sub-component is responsible for a simulation-based evaluation of a corrective measure's impact [13].

All components which are represented by micro services [70] are integrated using a manufacturing service bus for their orchestration. Integration is done by three well-known integration and communication patterns: publish-subscribe, workflow-based integration, event- and message-driven communication (Schel et al., 2018).

### Benefits of the Framework

The planning landscape at Infineon Technologies AG is built according the SCOR model and specified for each sub-process as depicted in Figure 4. Further explanations can be found in [76]. The developed framework is applied inside this planning landscape. The monitoring approach is used in Demand Planning where critical events are identified by means of the introduced control charts. For the monitored variable, the forecasts regarding the expected order quantities are used, which are extracted right before entering the planning of Operational Demand. This external information, sent by the customers, is one of the main inputs for the planning processes and therefore a major influencing factor in the system. Critical events in this context could for example be changes in the demand quantity or changes in the desired delivery date by the customer. Another

critical event would be the appearance of a totally new order, which would make an adaption of the planned scenario necessary. The detection of a relevant deviation triggers a process in the system for production management which automatically adjusts the production to the new situation by changing parameters via the ADC. To ensure automation this will mainly be performed by applying short-term decisions such as changing the priority of specific lots resp. changing the current flow factor of the production of a specific product.

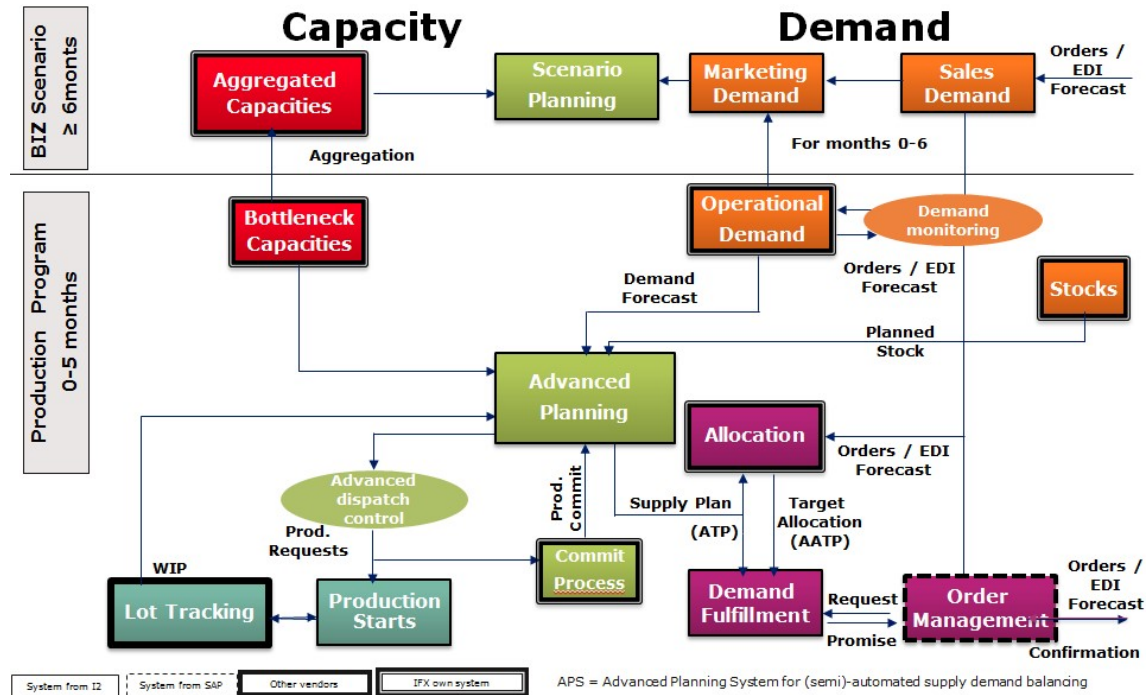


Figure 4: Planning landscape at Infineon Technologies AG

The applied concept allows for a reduction of the impact of unpredictable events on the supply chain's performance. First, this enables a faster reaction to events due to the elimination of media discontinuities and manual diligence tasks. Deviations to the as-is planning situation are recognized as soon as they occur. Second, the false alarm rate is reduced as a result of the consideration of the human behavior in the control charts. Third, on-time delivery is increased due to a possibility of intervention to production by applying corrective measures in production control. Fourth, inventory levels can be reduced through a higher target achievement of the delivery reliability. Summarized, these benefits lead to a more flexible and automated planning and production process and an adequate reaction to demand changes and customer requests.

### About Authors

**Dennis Bauer** studied Mechanical Engineering at the Baden-Wuerttemberg Cooperative State University Mosbach (Bachelor of Science 2012) and Production Engineering at the University of Stuttgart (Master of Science 2015). Since 2015 he is a research fellow and project manager at the Fraunhofer Institute for Manufacturing Engineering and Automation IPA in the department of digital tools for manufacturing.

**Andreas Schlereth** studied Industrial Engineering at the Karlsruhe Institute of Technology KIT (Bachelor of Science 2014 and Master of Science 2018). Since 2018, he works as a research fellow at the Fraunhofer Institute for Manufacturing Engineering and Automation IPA in the department of digital tools for manufacturing.

**Florian Maier** is a research assistant at the Fraunhofer Institute for Manufacturing Engineering and Automation IPA since 2009. In the "IT Applications and Services in Production" group of the digital tools for manufacturing department, he is responsible for IT and IoT structures in production, particularly with regard to the integration of industrial 4.0 concepts, work preparation, post-processing of data and escalation systems, as well as components for data visualization and interaction between man and machine. After he successfully completed his studies in business informatics at Furtwangen University of Applied Sciences. In addition to his work at the Fraunhofer IPA, he holds a Master of Business Engineering degree from the University of Stuttgart.

**Thomas Ponsignon** works as a Senior Staff Engineer at Infineon Technologies in the Corporate Supply Chain organization. He obtained Master's degrees in Industrial Engineering in 2006 from the EPF-Ecole d'Ingénieurs, Sceaux, France and the University of Applied Sciences, Munich, Germany and a Ph.D. degree in Mathematics and Computer Science in 2012 from the University of Hagen, Germany.

**Bernd Waschneck** works in the Frontend Manufacturing Analytics department at Infineon Technologies AG. He obtained his Master's degree in Physics in 2014 from LMU Munich.

## 2.6 MES Challenges

Authors: Critical Manufacturing; <https://www.criticalmanufacturing.com/en/homepage>

*Industry 4.0 is the vision of future manufacturing that allows not only lower cost, higher quality, and faster processing – but all of that with personalized products. Its a marketplace where smart products and smart equipment interact autonomously for dynamic optimization on the fly. For most companies, it's a concept and the way to implement it is not clear. While testing out new technologies and creating a new approach to production and supply chain operations, there is one critical foundation piece companies must lay, and that is a new Manufacturing Execution System (MES) that is Industry 4.0 ready. Recently successful manufacturing, particularly as it has gone global, has relied heavily on Manufacturing Execution Systems (MES) - also sometimes called Manufacturing Operations Management (MOM). This multi-faceted software for production plants has been a pivotal enabler for the performance, quality and agility manufacturing leaders have achieved. Unfortunately, not every MES in use today is capable of supporting Industry 4.0. In the Table 1 are compared traditional MES with the New MES..*

Traditional MES	New MES (Industry4.0-ready)
Fixed model of a plant, fixed screens to view	IIoT event, location processing for context, augmented reality, mobile interfaces
On-premise, some intelligence for floor	Cloud-based; advanced analysis inside plant context for big data
Linear, fixed model of assets and processes	Binding to connect in real-time with context resolution
High-touch integration to automation and ERP	Dynamic execution of processes and status in a shop floor marketplace
Plant-context workflow	Service-oriented, modular visibility across operations

Table 1: Industry 4.0 requires characteristics that many current MES do not have, so manufacturing IT must evaluate carefully.

### The Future of MES/MOM Under the Industry 4.0 Transition

Will MES continue to play an essential role in the future manufacturing enterprise's IT landscape and why? The answer to these two pivotal questions resides on the fact that MES sits at the critical point where revenue-generating products come into being. MES already handles rapidly flowing streams of disparate data and turns it into useful information in a near-real-time fashion. The influx of Cyber-physical system (CPS) and Cyber-physical production system (CPPS) data will require this and more. The primary functions of MES are a critical foundation around which manufacturers can build the Industry 4.0 application structure. As it is today, MES provides critical information both within the production environment and to the supply chain, customer service, product development and management teams. Industry 4.0 will not be fully implemented overnight, so there will be a transition period.

Over the longer term in Industry 4.0 situations, MES will play several roles.

- sit at the center of the smart supply chain
- monitor, and as needed, coordinate the CPS-CPPS marketplace
- incorporate CPS and CPPS data into off-line compliance and quality activities



- act as the stand-in for any products, materials, or equipment that are not fully CPS enabled
- be the essential coordinator for both the horizontal and vertical integration of Industry 4.0
- sit at the center of the product lifecycle
- deliver aggregated information for customer service and other eco-system activities

### Characteristics of MES of the Future

Manufacturing Execution Systems have been pivotal in the performance, quality and agility needed for the challenges created by globalized manufacturing business and will most likely continue to be. However, a completely new generation is required to cope with the new challenges created by Industry 4.0. The following are the five main pillars these systems shall consider in order to support Industry 4.0 effectively.

1. **Decentralization:** Industry 4.0 is inherently a decentralized system, with intelligence in independent entities. Smart materials and products (CPS) are service consumers and smart equipment and plants (CPPS) are service providers. CPS and CPPS are not physically coupled; rather, a dispatching operation delivers logical binding between a material to be processed and a resource to process the material. Since each product in the Industry 4.0 vision may be unique, it will be very difficult to centralize or optimize shop floor operations in the traditional way. This dynamic marketplace of CPS and CPPS means that rather than hold a single unifying model, the MES needs context resolution possibilities. This allows a product that requires a certain service at a certain step to combine the flow of its product category to be adapted, or unique to its specific context. Going one step further, the smart product may hold the recipe needed at a given processing step. When negotiating with the smart resource, it will transfer the recipe to the resource so that it can perform its unique transformation process.

So the CPS and CPPS have their own intelligence. As an example, a smart product CPS knows its state, its position, its history, its target product and its flow alternatives. Likewise, a smart resource or CPPS will know its state, its history, its maintenance plan, its capacity, its range of possible configurations and setups, etc. What this means is that a smart product or CPS has the capability to identify itself, providing its position and state to a physically centralized system. MES decentralization needs to be logical, not necessarily physical. In fact, with cloud computing, it's even arguable whether such a system can be considered physically centralized. What is critical is that the logical decentralization must exist. So the MES may still be one centralized application, but it acts decentralized with agents or objects representing the shop-floor entities.

2. **Vertical integration:** MES has always been most effective when integrated into the ERP systems 'above' it as well as the monitoring, control and actual production processes 'below' it. This vertical integration is an enabler for the orchestration of business processes that may be simple or complex but nearly always require multiple layers and groups to be involved. Business processes for compliance, quality, logistics, engineering, sales or operations all have components inside the plant as well as others that reside beyond the factory that are crucial to a business process being executed effectively.

In Industry 4.0, CPS and CPPS communications create new data flows to integrate. For example, a CPS or smart product may know that it needs to stop a lot or collect a measurement variable. The layer above then specifies and checks whether the outcome is correct. This is where statistical process control (SPC) rules reside, and if it's not within limits, it might open a corrective and preventive action (CAPA). All of that activity will be in a layer above the CPS in controls and/or MES.

Within the plant, the MES will need to aggregate and put these additional data flows into context. Vertical integration of these autonomous entities is critical as they could otherwise make decisions independent of the rules and best practices for the factory or company. MES



providers must continue to expand the product capabilities to ensure that all plant activities are visible, coordinated, managed and accurately measured. Only then can the enterprise systems respond effectively.

3. **Horizontal integration** In the Industry 4.0 final report, horizontal integration focuses on supply chain status communication among facilities and trading partners. Horizontal integration enables the smart supply chain or network to be transparent so status is always visible. What horizontal integration requires is service-oriented ways of alerting the rest of the information system to the information available.

However, that is not likely to all come from the CPS and CPPS directly. In Industry 4.0 the MES must be truly modular and interoperable so that all functions or services can be consumed by CPS smart materials, CPPS smart equipment or any other shop-floor entity. As an example, a typical maintenance management process, often centralized, could consist of a series of services that each piece of equipment might use. Horizontal integration may also extend from the plant systems of suppliers to connect into the plant systems of their customers. With high tech's extensive outsourcing and supplier base, this is a vision these industries could benefit from greatly.

4. **Connectivity, sensing and mobile:** Advanced manufacturing environments have had highly integrated connectivity for a long time. As an example, some of the more sophisticated semiconductor facilities have RFID transponders in the material containers and the equipment has bidirectional communication through interfaces, exposing readings from sensors, alarms or reports or allowing recipes to be externally selected or downloaded. Now, Industry 4.0 is creating a true democratization of such connectivity, allowing it to be widespread in manufacturing facilities of different sophistication levels. Two elements contribute:

- (a) The IoT, in the industrial world called IIoT (Industrial Internet of Things) translates into very low cost hardware and lean OS (such as Windows 10 IoT running on a Raspberry Pi), allowing true connectivity with equipment not requiring heavy systems and interfaces.
- (b) Increasingly affordable passive identification and location tags allow all shop-floor resources (CPS and CPPS) to hold their positioning coordinates. The MES needs logically autonomous entities to store this location data and show it in real-time in interactive maps.

On the operational MES front, connectivity and mobile combined will allow more adaptable interfaces. MES will consist of different apps, making a reality the vision of getting to a piece of equipment, downloading and later using an app specifically built to operate that equipment.

The same combination of mobile devices with the increase in reliable and inexpensive positioning systems will also allow the representation of real time positioning in 3D maps, opening the door to augmented reality scenarios. A person with augmented reality can walk around and get immediate identification of items and be pointed to their location. Maintenance operations can benefit particularly.

5. **Cloud computing and advanced analysis:** The Smart Factory vision of Industry 4.0 requires achieving a holistic view of manufacturing operations. Clearly this can only happen by integrating data from several different sources rapidly and flexibly. This suggests the MES of the future must also leverage cloud computing and advanced analytics.

While many MES have manufacturing intelligence components today, this must expand to better accommodate the diversity and volume of big data. Both CPS and CPPS will generate huge amounts of data, which needs to be stored and processed. Advanced analytics are then needed to fully understand the performance of the manufacturing processes, quality of products and supply chain optimization. Analytics will also help by identifying inefficiencies

based on historical data and pointing staff to corrective or preventive actions for those areas. Future MES must accommodate both:

- (a) Advanced offline analysis using very sophisticated statistical process models. These will need to be both in structured data, generally residing in a relational database or in data warehouse cubes, and in unstructured data, which is very difficult to analyze with traditional tools.
- (b) “Real-time” analysis to trigger actions in the plant as quickly as possible, even before data is stored. This needs techniques such as “in-memory” and complex event processing.

Cloud computing is the obvious infrastructure for the speed and agility suggested by Industry 4.0. On-premises systems have far more limited ability to expand, change and respond, and would result in undue costs. Manufacturing data analysis is an area where some leading manufacturers are already starting to leverage the cloud.

## 2.7 Machine Learning Algorithms in the Semiconductor Industry

Author: Felix Kamhuber; Fraunhofer Austria Research GmbH

Author: Alessandro Chiancone; Know-Center GmbH Research Center for Data-Driven Business  
Big Data Analytics

*This contribution features a concept solution for fast systematic Root Cause Analytics (RCA) on the feedforward product variability. The solution is induced by several critical upstream Redistribution Layer (RDL) assembly processes in order to reduce and control significant impact of process variation of contact resistance as well as the variation of copper contrast at the Automatic Optical Inspection (AOI) device. The overall target is the reduction of the Overkill Impact, representing one of the key output control variables. Statistical Process Control (SPC) and Virtual Metrology (VM) are applied to detect faulty wafers. To gain this conceptual solution for RCA, the following core steps should be taken into consideration, namely i) Data Input Preparation and Visualization, ii) Data Preprocessing and Transformation followed by iii) a comprehensive Root-Cause Analysis carried out by supervised Machine Learning Algorithms covering two Classification approaches and one Regression method. The current cost-intensive manual performed review on wafers is unsystematic and time consuming. The core research question is affected by the lack of input variables and thus it explores whether a valid and reliable correlation can be identified using available output data only provided by the AOI metrology processes measured versus the tracked data. This contribution discusses the aforementioned question and presents the key findings including assessment and validation results.*

### Introduction

The introduction covers the use-case setting, providing insights into as-is state and demands, and formulates the problem statement and research methodology.

### Use-Case Setting

The international semiconductor manufacturer described herein packages single Chips (dies) to one package (System in Package) and outlines conductors to single pins. These pins allow the package to be placed on a PCB (Printed Circuit Board). Single Chips are cut from the supplied wafer or supplied format, placed on a reconstructed wafer, including enough space to outline the conduction routings to the single pins. After creating new routings in often more than one layer, the conductivity of this routings has to be ensured. The conductivity can be measured either electrically via contact resistance or optically based on inspection of copper colorization of the routings. The optical inspection is based on a reference image (perfect example, Gold Standard), marks any deviation and leads to an alarm. Every lot with excessive number of alarms causes to set this lot on hold. Every lot on hold slows down the production process and at least costs money. Every alarm requires manual revision to clarify the correctness of the alarm. Significant problems require engineering efforts to adopt the process to keep a high yield. In future, CR (Contact Resistance) and AOI (Automated Optical Inspection) should be more stable by reducing false alarms. The manual revision should be engaged with real defects only and false alarms should be reduced.

### Problem Statement

One of the main goals is to improve AOI error detection rate in order to reduce the Overkill Impact, representing the redundant checks due to false alarms. To this end, “FirstPass Units” refers to a high percentage of acceptable wafer units for each wafer and “Pass Review Units” to the few wafer units that are reviewed for the 2nd time (see Figure 1). On this basis, four output variable relations can be derived between defects and units metrics on wafer unit level as follows:

1. FirstPass Units + PassReview Units + Reject Units = Total Inspected Units
2. Defects Total – Defects Good = Total Defects set as fail Reject Units
3. Defects Total PassReview Units (units with all Defects set as Good) + Reject Units (units with at least one defect set as fail)
4. Defects Good PassReview Units (units with all Defects set as Good)

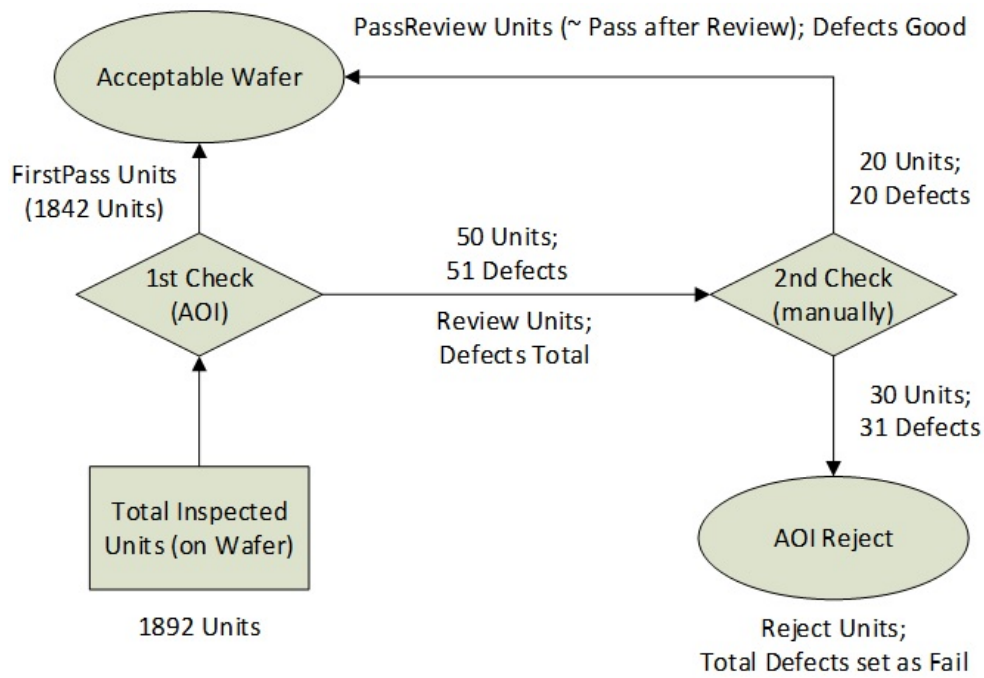


Figure 1: Relations between defects and units metrics

Within each data set, three output labels with correlation to their respective unit metrics are provided as follows: Overkill Impact, Defect Efficiency, Yield Impact (see Table 1)

AOI Output Labels	Unit Metrics
Overkill Impact	Defects Good / (FirstPass Units + Defects Good + Reject Units)
Defect Efficiency	(Defects Total – Defects Good) / Defects Total
Yield Impact	1 – Units Reject / (FirstPass Units + Defects Good + Reject Units)

Table 1: Correlation output labels - Unit Metrics

### *Research Methodology*

Predictive data analytics project lifecycle – as defined by the Cross Industry Standard Process for Data Mining (CRISP-DM) considers six core phases in a predictive data project lifecycle, namely 1) Business understanding, 2) data understanding, 3) data preparation, 4) modeling, 5) evaluation and 6) deployment [128].

The methodological approach applied on this use case is based on the CRISP-DM model, focusing on the first five phases. The modelling phase is based on a first order root cause analysis derived by the output variables that can be, according to the project specification, summarized to the AOI output labels. Modelling based on a second order RCA is – due to the lack of available critical input variables – not possible.

Due to the lack of exploitable information, besides the given output labels and the general wafer sequence information, the authors decided to use R Studio [168] [176] for the Data Preprocessing steps and to employ following classification algorithms, namely Random Forests and Artificial Neural Networks (ANNs). The latter ones are well known for their capability to model nonlinear relationships and/or interactions [36]. The former ones provide an improvement over bagged decision trees, whereby bagging stands for a general-purpose procedure for reducing the variance of a statistical learning method [95]. The aforementioned algorithms were complemented by a Non Negative Least Squares regression approach in Matlab [140], being able to compete them with each other, and finally assessed by the use-case partner.

### *Definitions and Literature Review*

Predictive modeling defines the process of developing a mathematical model, based on patterns from historical data that generates an accurate prediction [128] [116]. These patterns are highly dependent on predictive signals in data sets. Several models may discover a signal to some degree of accuracy, however, the credibility of model building has weakened especially as the window of data access and analysis has been widened. Common culprits include, next to inadequate pre-processing and model validation, especially over-fitting the model to existing (historical) data [116].

Machine Learning, being defined as an automated process extracting patterns from historical data, is applied in the fourth phase (modeling) [128]. Although Machine Learning (ML) algorithms can be successfully applied in industry for general job-shop scheduling problems or the prediction and control of product quality within a mechanical process, there are, according to Stricker [209], several requirements acting as key enabler for ML within industrial applications, namely: (i) applications with limited scope in terms of dimensions and actions, (ii) availability of a lot of training data (ML approaches are intensively data-driven) and (iii) complex environments that can hardly be described in detail. Toly Chen [36] [35] proposes an innovative approach based on multi-layer ANNs for modeling and prediction of the yield learning process of a semiconductor product. Lin and Chen [127] propose simulation-based adaptive optimization strategies to deal with complex semiconductor manufacturing systems in terms of hybrid flow shop scheduling. Kim [110] proposes machine learning-based detection for faulty wafers in semiconductor manufacturing processes, with focus on Dimensionality reduction.

Based on the above-mentioned examples, the range of industrial successful applications is demonstrated. Lingitz describes in [128] the ten most important variables of the models resulting from a sensitivity analysis. Within our research we want to give an answer on the question, if the application of ML algorithms provides to be useful without the use of additional input features, besides the given wafer sequence itself.

### *Data Preprocessing and Visualization*

Data pre-processing, covering the second and third phase within CRISP-DM, includes the following described tasks being processed, based on a predefined data structure. For a better understanding

of the data and for the generation of the only feature (besides the already given output labels) the sequence has to be constructed and visualized. The numerical sequence is gathered based on (unique) time-stamp information, whereby the time-stamps are provided on a very detailed level (accuracy: seconds) to avoid redundancies. As a control mechanism, a general ‘known path-order’ is used for sequence control (natural ordering), to automatically detect and correct data outliers. The calculation of the key performance indicators (output labels) has to be carried out by the provided input metrics (see table 1) within the specific data set. The creation of a path profile provides an overview, how many different unique wafer sequence paths (per data-set on approximately 1000 wafer IDs) exist. It is revealed that depending on each unique data set, 10 – 20 paths (on station level) exist. Without taking into consideration the detailed wafer sequence, one must distinguish between:

1. (Simplified) Sequence information on station level
2. (Sophisticated) Sequence information on a more detailed level: mapping station – sub equipment

This contribution deals with both levels of sequence information. The final assessment from the Use-Case partner is based on the second level, being much more complex in comparison with the first one. A wafer sequence consists of around 20 – 25 stations in total (per sequence), while there can exist several sub equipment to one station, leading to a complexity of approximately 160 combinations station – sub equipment. The overview provides the advantage of neglecting outlier sequences that have occurred below a certain percentage threshold level, e.g. 1%. The latter threshold is used in combination with classification algorithms (described in the next section) and leads to a 20% reduction of total observations on the given data set.

A network visualization of the sample data based on the detailed level was created (see Figure 2) offering an interactive and descriptive view on the combination of the used stations (green, with text) and equipment (blue, with text) of each Wafer-ID. The visualization was created in R using the ‘visNetwork’ package [4]. It shows the sequence of a wafer through the fab, whereby used stations are represented in green boxes, while used sub-equipment on the corresponding station are figured in blue ones.

- Referring to the figure below it is obvious that on a station a wafer can pass through more than one sub-equipment, for instance Plating process: the wafer has to pass 3 sub-equipment serially
- Each wafer ID can be visualized separately

Another step of the pre-processing phase includes the handling of ‘datafound’ positions: All of these positions are analyzed on each station, and if ‘datafound’ entries are found, they are replaced by the most often used sub equipment on this machine. If there are only ‘datafound’ entries in the whole data set, they cannot be replaced at this point. As preparation in order to apply classification algorithms, the sequence was simplified into a binary sequence representation acting as a feature for the classification.

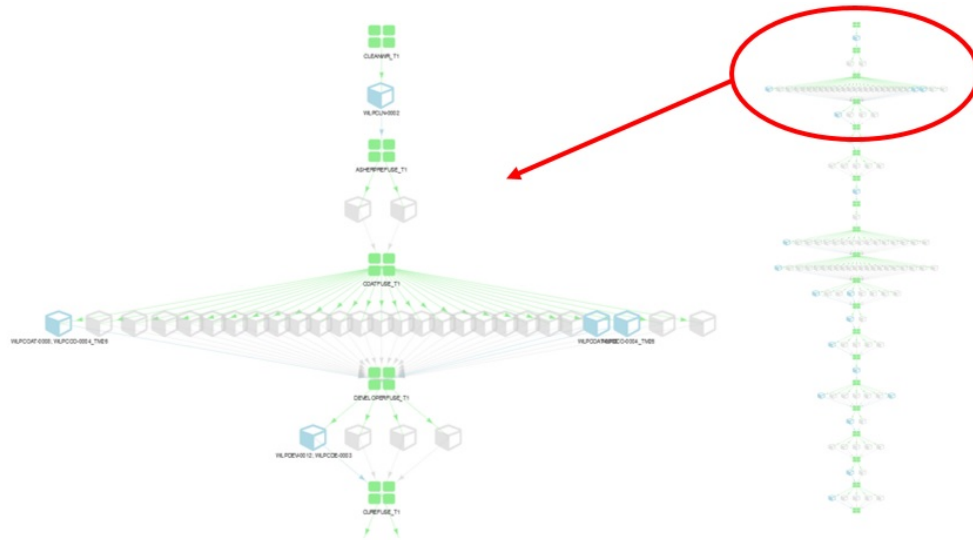


Figure 2: Network visualization (left: detailed view; right: full network view)

### Classification Algorithms

In order to solve the classification problem, Random Forests are the preferred choice over Decision Trees. When it comes to apply Tree-Based Methods, they are built of a number of decision trees bootstrapped in training samples, with one difference as follows: Each time a split in a tree is considered, a random sample of  $m$  predictors is chosen as split candidates from the full set of  $p$  predictors. Thus, the algorithm is not allowed to consider a majority of the available predictors, making it more robust against one single very strong predictor [95].

In comparison with an intrinsically linear tree-based method, one nonlinear classification model (an ANN, with different options and settings) was applied on the classification problem, facing the same problem like its regression counterpart, the significant potential for over-fitting [116, p. 335].

#### *Random Forests*

The classification is based on sequence and wafer-level output information. The binary-coded stations serve as input labels, with the corresponding Overkill Impact level as output label. Based on this binary coded sequence representation Random Forests could be successfully applied. The output label corresponds to all available (approximately 20) input stations (of the corresponding wafer sequence), because the specific influence (maintenance intervals, condition, etc.) of each station is considered to be unknown. The data sets are split into 2/3 training and 1/3 test data, for making predictions.

The Variable Importance Plot shows a dot chart of variable importance as measured by a Random Forest (see Figure 3). Variable importance is computed using the mean decrease in the Gini index, that is defined by a measure of total variance across the  $K$  classes, and expressed in relation to the maximum [95]. The left side, for instance, shows the mean decrease of the accuracy in relation to certain input labels. According to Figure 3, most of the variables (stations) have very few or no impact on the accuracy of the created random forest (according to the analyzed data set). The variables with the greatest impact are listed first.



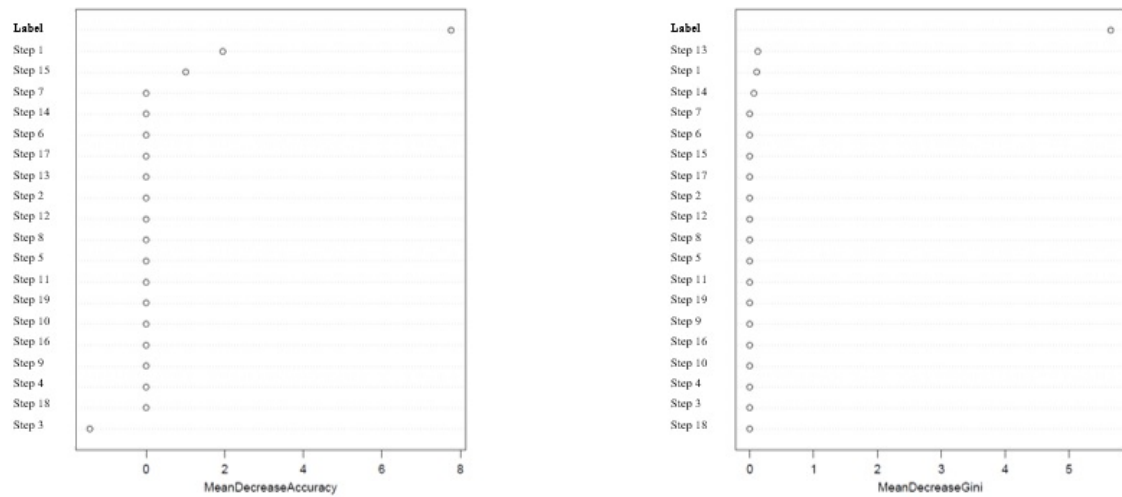


Figure 3: Variable importance plot (random forest)

Predictions based on the 3 predefined categorical threshold levels (“Low”, “Mid”, “High”) seem to provide promising results with a high accuracy, but a deeper view into these results reveals that > 95% of the wafer IDs have the “Low” threshold level. Based on this conclusion ANNs are applied to have a closer look on a numerical based prediction of the precise Overkill Impact value.

### *Artificial Neural Networks*

Based on the binary sequence representation ANNs have been created and tested using the ‘neuralnet’ [61] package from R Studio, relating to a (weak) formula fm, based on all binary-coded stations, and the default algorithmic options (see Figure 4). Besides the pre-processed sequence information, no evaluable network information network can be provided that could be processed on a higher machine level, because all stations can have an impact on the output. As only output label, the Overkill Impact was used here.

```
#Resilient Backpropogation algorithm (RPROP+) -> learningrate.limit, learningrate.factor
nn = neuralnet(fm, data = predictors_matrix_train, hidden = 1,
  threshold = 0.01,
  learningrate.limit = NULL,
  learningrate.factor =
    list(minus = 0.5, plus = 1.2),
  algorithm = "rprop+")
```

Figure 4: Default ann (RPROP+) options

Besides the default options for a resilient backpropagation algorithm (see Figure 4), different settings with multiple hidden layers have been tested (see Table 2) on the data set.

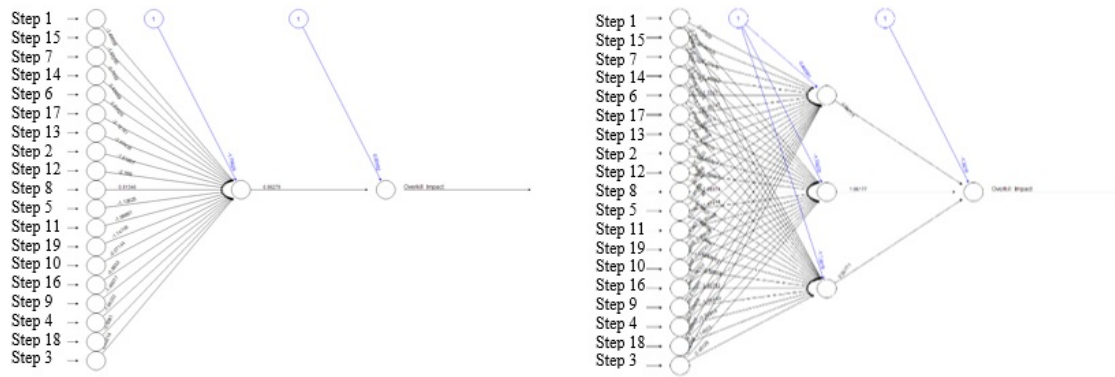


Figure 5: Resilient backpropagation algorithm (RPROP+): 1 Hidden Layer (left), 3 Hidden Layers (right)

While decision trees and random forests accept categorical variables (a binary structure was implemented before), ANNs only accept numerical inputs. The range of the actual values is  $[0, 0.09]$ , while the range of the predict values varies between  $[0.001, 0.0038]$ . It seems that the prediction is not able to predict 0 values, when there actual is no overkill given. The full data set includes 2400 entities, while the reduced data-set includes 1900 observations.

Ann configuration	Accuracy
'Default' (rprop+; Hidden Layers: 1)	3.4% (21/614)
Hidden Layers: 3	3.7% (23/614)
Hidden Layers: 5	2.0% (12/614)
Hidden Layers: 2 (Neurons: 10, 3)	1.1% (7/614)

Table 2: Accuracy of different ANN configurations

Finally, ANNs are used for Overkill Impact predictions. Those predictions have a low quality ( $< 5\%$  accuracy), dependent on the data-set being used and the used tolerance levels. For the comparison in table 2, the tolerance level was set to 25% Higher tolerance levels (about 30-40% deviation from the actual overkill value) in combination with a bigger (merged) data-set resulted generally in a higher accuracy (up to 20%). Using different algorithmic options and data sets, a satisfying accuracy rate has not been achieved. One main reason for the low accuracy rate shows figure 7 indicating that the ANN itself has especially difficulties in predicting actual outlier values (actual values above 0.01 and 0 itself).

#### *Non Negative Least Squares (NNLS)*

As a complementary approach, the problem has been formulated in terms of regression assigning a vector  $x$  to each wafer with a binary variable per station with 0 value, if the wafer did not pass through the combination station-equipment and 1 otherwise. Output measures  $y$  are then regarded as target variable (response variable) as in:

$$y = f(x), x \in \{1, 0\}^n$$

Where variable selection, the estimation of the link function is not the direct task of this use case. In order to obtain meaningful results, assumptions have been made that each combination station-sub equipment can have zero impact on the quality of the wafer or it can have a negative impact. This assumption is crucial since a direct consequence is that errors made in station  $q$  will persist until the evaluation step. Notably, the model allows no cancellation effect in the following

stations. Furthermore, wafers are assumed to be in perfect condition when they enter the pipeline. Linear regression cannot be considered since the estimated coefficient can assume negative values. Non-negative least squares has been considered to solve this problem:

$\operatorname{argmin}_x \|Ax - y\|$  subject to  $x \geq 0$  which is equal to a quadratic programming problem

$$\operatorname{argmin}_{x \geq 0} \left( \frac{1}{2} x^T Q x + c^T x \right) \text{ where } Q = A^T A \text{ and } c = -A^T y$$

This approach is linear [119], i.e. justified by the intrinsic design of the network and the underlying assumptions. Non-linear techniques have been nonetheless used to assess if better prediction accuracy can be achieved. Nonlinear techniques such as random forests or support vector machines tend to rapidly overfit having in general worse performances and lower interpretability. Datasets from 10 different and independent periods (1 period = 1 production week) have been studied. The datasets contain different amount of entries with an average of 1000 points. Each dataset has been analyzed independently since the variable importance can change overtime and is assumed constant over a single dataset.

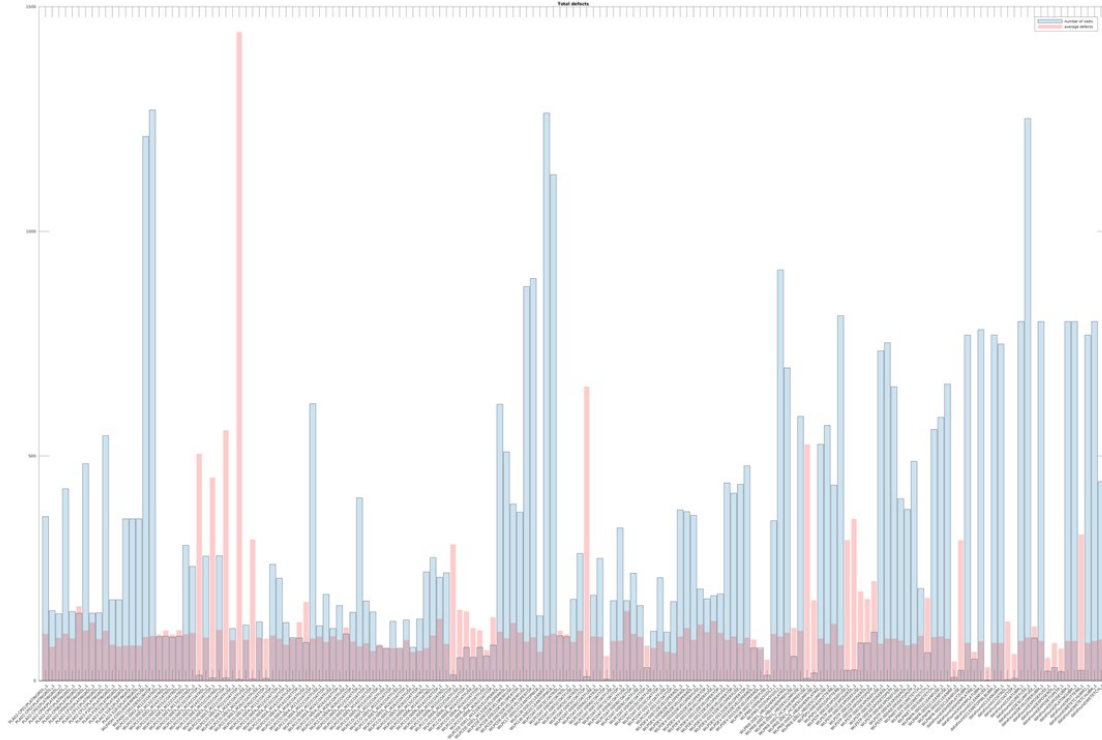


Figure 6: AVG. Total defects (red) over number of visits (blue) per sub-equipment

Diagnostic and prediction results have been provided to the partner company for evaluation and comparison purposes. Different data sets show a different topology of the problem, i.e. number of stations and number of visits per station. As exemplified in Figure 6, it is evident that the number of visits is not uniformly distributed. In order to obtain significant results many (more) wafers are needed to explore the network, if no further assumptions are used to constrain the problem. Once the solution A is computed, the coefficient relative to each single station shows the impact on the output measure, i.e. the larger the coefficient the larger is the negative impact of the station on the process. Figure 7 shows the prediction of the Overkill Impact, as expected, the linear solution is

not capable of grasping the big fluctuations present in the data. Evidently, this is because of the fact that even if two wafers follow the path and visit the same machine (i.e.  $x_i = x_j$ ), the Overkill Impact factor can fluctuate significantly. The reason for this behavior has different possible explanations:

- Some key features are not accounted in the data
- The initial full integrity of the wafers cannot be assumed
- An error occurred in the station in a time in between the two wafers, i.e. the assumption of stability of the process internal status does not hold (i.e. the vector  $A$  is changing over time)

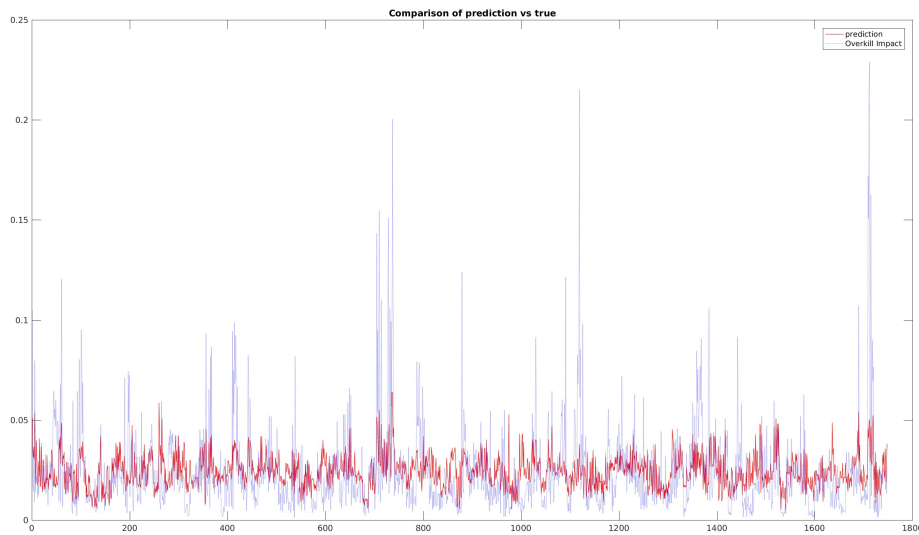


Figure 7: Prediction of the overkill impact factor (red) based on the solution obtained with non-negative least squares (blue)

### Assessment Results

The evaluation by the company was performed based on the NNLS method on wafer level, because only the last approach was applied on the deeper level (combination station-sub equipment). This assessment includes the evaluation of diagnostic results on the one side, and predictive results on the other side, applied to wafers of high impact. This implies that only the impact coefficients of the wafer used pairs station-sub equipment are taken with the respective traced root cause information on the quality of the internal quality deviation records. Additionally an impact coefficients weight criteria was applied on the True Positive Results. The applied weighting set comprises  $w=1, 0.75, 0.50$  according to the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> largest coefficient.

Smart Data-Set	Overkill Impact (> 15%)				Yieldloss Impact (> 1%)			
	Critical Events (wafers)	Traced RC record	True Positive	Recall (TPR)	Critical Events (wafers)	Traced RC record	True Positive	Recall
1	9	2	0	0%	8	5	0.5	10%
2	11	3	0	0%	8	4	0	0%
3	12	2	0	0%	14	8	0	0%
4	8	4	1	25%	7	5	0.5	10%
5	3	0	0	-	7	2	0	0%
6	1	0	0	-	7	3	0	0%
7	0	0	0	-	5	0	0	0%
8	1	1	0	0%	7	2	0	-
9	3	0	0	-	6	0	0	0%
10	4	0	0	-	11	5	0.5	-
<b>Total Events</b>	<b>52</b>	<b>12</b>	<b>1</b>		<b>80</b>	<b>34</b>	<b>1.5</b>	<b>10%</b>
<b>% Used Events/ Overall Recall</b>		<b>23%</b>	<b>8.3%</b>			<b>43%</b>	<b>4.4%</b>	

Table 3: Comprehensive assessment of diagnostic results

The quantitative assessment of the predictive results can be applied on all 3 output labels (Defect Inefficiency, Overkill Impact, Yieldloss Impact) yielding in a confusion matrix – consisting of the counted True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) – and values for Accuracy and Recall.

TP	0	0	0	0	0	0	0	0	0	0
TN	1729	1423	1223	774	1090	1618	1030	1367	598	1248
FP	0	0	0	0	2	0	0	0	3	0
FN	18	25	30	42	51	81	22	29	21	22
Accuracy	99.0%	98.3%	97.6%	94.5%	95.4%	95.2%	97.9%	97.9%	96.1%	98.3%
Recall (TPR)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table 4: Assessment of predictive results on output label ,Defect inefficiency‘

TP	0	0	0	0	0	0	0	0	0	0
TN	1738	1437	1241	808	1140	1698	1052	1395	619	1266
FP	0	0	0	0	0	0	0	0	0	0
FN	9	11	12	8	3	1	0	1	3	4
Accuracy	99.5%	99.2%	99.0%	99.0%	99.7%	99.9%	100.0%	99.9%	99.5%	99.7%
Recall (TPR)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-	0.0%	0.0%	0.0%

Table 5: Assessment of predictive results on output label ,Overkill impact‘

TP	0	0	1	0	0	0	0	0	0	0
TN	1745	1448	1250	814	1141	1699	1052	1393	622	1269
FP	0	0	0	0	0	0	0	0	0	0
FN	2	0	2	2	2	0	0	3	0	1
Accuracy	99.9%	100.0%	99.8%	99.8%	99.8%	100.0%	100.0%	99.8%	100.0%	99.9%
Recall (TPR)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-	0.0%	0.0%	0.0%

Table 6: Assessment of predictive results on output label ,Yieldloss impact‘

The quantitative assessment of the diagnostic and predictive results (Tables 3 – 6) yields a negligible Recall result, but a high accuracy level (> 99%). As shown in Table 3, the NNLS method has problems in predicting outliers and the very few traced root cause events correctly. This can be

justified by the complexity of the network, and the small size of the received data sets. Furthermore, the prediction is not based on the threshold values that are used by the company in the final assessment phase. Accuracy and Recall would be higher, if the assessment would be based only on wafer level, not in combination with a set threshold for a given output label – see Table 7. Obviously, the predictions for Wafers 1 – 3 are also correct (orange marked cells), however, the corresponding threshold (level defined by the company) for Overkill Impact was not achieved for these Wafer ID's. Thus, the recall was not recorded at these points. This table shows that, in total, 15 critical (wafer) events are logged (as stated in Table 3), whereby for 6 events no root cause is recorded. Compared over all 10 assessed data sets 36 out of total 132 events ( 27%) can be assigned a traced root cause, implying that for all other cases the traced root cause is unknown.

Wafer ID	Overkill Impact	Yieldloss Impact	Reported Quality Issue	Traced Root Cause	Overkill Impact			Yieldloss Impact		
					First (1)	Second (0.75)	Third (0.5)	First (1)	Second (0.75)	Third (0.5)
Wafer 1	0.26%	1.11%	Confidential	M_1 E_T1	M_1 E_T1	M_2 E_T1	M_3 E_T1	M_3 E_T1	M_8 E_T1	M_2 E_T1
Wafer 2	0.69%	1.48%	Confidential	M_1 E_T1	M_1 E_T1	M_2 E_T1	M_3 E_T1	M_3 E_T1	M_8 E_T1	M_2 E_T1
Wafer 3	0.90%	1.37%	Confidential	M_1 E_T1	M_1 E_T1	M_2 E_T1	M_3 E_T1	M_3 E_T1	M_8 E_T1	M_2 E_T1
Wafer 4	15.08%	0.11%	Confidential	M_18 E_T1	M_1 E_T1	M_2 E_T1	M_3 E_T1	M_3 E_T1	M_8 E_T1	M_2 E_T1
Wafer 5	15.98%	0.11%	Confidential	M_18 E_T1	M_1 E_T1	M_2 E_T1	M_3 E_T1	M_3 E_T1	M_8 E_T1	M_2 E_T1
Wafer 6	16.60%	0.05%	Confidential	M_1 E_T1	M_1 E_T1	M_14 E_T1	M_3 E_T1	M_3 E_T1	M_8 E_T1	M_16 E_T1
Wafer 7	8.84%	2.01%	Confidential	M_19 E_T1	M_2 E_T1	M_3 E_T1	M_4 E_T1	M_3 E_T1	M_8 E_T1	M_12 E_T1
Wafer 8	3.59%	1.11%	Confidential	M_20 E_T1	M_3 E_T1	M_15 E_T1	M_4 E_T1	M_3 E_T1	M_8 E_T1	M_20 E_T1
Wafer 9	1.32%	1.00%	No root cause record		M_2 E_T1	M_3 E_T1	M_4 E_T1	M_3 E_T1	M_8 E_T1	M_12 E_T1
Wafer 10	22.94%	0.32%	No root cause record		M_5 E_T1	M_2 E_T1	M_4 E_T1	M_2 E_T1	M_9 E_T1	M_10 E_T1
Wafer 11	18.83%	0.16%	No root cause record		M_5 E_T1	M_2 E_T1	M_4 E_T1	M_2 E_T1	M_9 E_T1	M_10 E_T1
Wafer 12	4.75%	9.49%	No root cause record		M_2 E_T1	M_6 E_T1	M_7 E_T1	M_2 E_T1	M_10 E_T1	M_10 E_T1
Wafer 13	18.93%	0.11%	No root cause record		M_5 E_T1	M_2 E_T1	M_4 E_T1	M_2 E_T1	M_10 E_T1	M_10 E_T1
Wafer 14	22.26%	0.16%	Confidential	M_20 E_T1	M_1 E_T1	M_13 E_T1	M_7 E_T1	M_17 E_T1	M_10 E_T1	M_10 E_T1
Wafer 15	39.21%	0.03%	No root cause record		M_2 E_T1	M_6 E_T1	M_7 E_T1	M_12 E_T1	M_2 E_T1	M_9 E_T1

Table 7: Assessment details (data-set 4)

**Conclusion** [119] Limitations of the applied approaches (Random Forests, Artificial Neural Networks, NNLS method) are due to different factors, namely i) topology of the network, ii) number of visits, iii) uncertainty of the input and iv) evaluations process. In real scenarios, factors that affect the quality of the solution are prevalently not always known in advance. It is, therefore, reasonable to evaluate, based on the data provided, whether the obtained result feature good stability and accuracy properties. The topology of the network in this use case is of crucial importance. If bottlenecks are present, then equally likely multiple solutions are found. How the network is visited, affects the speed of convergence of the algorithms, i.e. if a node is never visited or just a few times the information is weak to draw statistically significant conclusion. A simulation has been performed assuming uniform transition probabilities and normally distributed errors around a given mean. The non-negative least squares in such conditions is capable of retrieving the mean of each single node enabling to point to stations that are more affecting the quality of the wafers. In case the distribution is heavy tail, in the extreme of a Cauchy distribution the algorithm could not retrieve the correct solution. This is pointing out the fact that in presence of great uncertainty, this problem cannot be solved with non-negative least squares and more assumptions should be made to constraint the solution. The uncertainty of the input arises when two wafers that visit the same stations have extremely different output measures, in this case there could be multiple explanations: first one of the input wafers is damaged (different charge, recipe and/or product type) and second the state of the system has changed (due to maintenance or staff) and an error occurred in at least one of the station visited. According to the evaluation of the results by the company, it is evident that for most of the cases the traced root cause is unknown.

The most significant conclusion of this methodological research on the network on two, hierarchically different views (station level on the one hand and the combination of station-sub equipment level on the other hand) reveals the answer to the initially postulated research question. In particular, the combination of very few available observations per data set (approximately 1000) in

combination with only few information about the topology of the complex network (the sequence of each wafer as only labelled input and binary coded as a combination of approximately 160 stations-sub equipment pairs) itself and the mentioned uncertainty of the input does not admit a significant reliable diagnostic analysis of the as-is situation and prediction. Further research should integrate predictions with critical process input variables ('features'), that cannot be delivered by the company at the moment due to the ongoing sensor hardware integration on site. The performed research studies should be repeated once, if the company is able to provide larger data sets ('Big Data') to solve the unresolved question, if the applied approaches would deliver better results on a larger test sample.

### Acknowledgments

This work and book contribution has been performed within the project Power Semiconductor and Electronics Manufacturing 4.0 (SemI40) under the grant agreement No 692466. The project is co-funded by grants from Austria, Germany, Italy, France, Portugal, and Electronic Component Systems for European Leadership Joint Undertaking (ECSEL JU) [187]. We also would like to thank our research partners within the SemI40 project for their support throughout the project, especially Alessandro Chiancone (KNOW center) and our main contact person from the company.

### About Authors

**Felix Kamhuber** finished his bachelor and master studies in 'Industrial Logistics' at the University of Leoben with emphasis on Information Technology and Automation in 2010 and is working as researcher since 2014 (PHD candidate) at Fraunhofer Austria Research GmbH in Vienna (Austria) with focus on the following research areas and topics: Simulation Based Optimization, Mathematical Optimization (e.g. Combinatorial Optimization), Operations Research in the context of production planning problems (e.g. Production Smoothing, Permutation Flow Shop Problems, Machine Assignment Problems), Data Science and Statistical Computing.

**Alessandro Chiancone** obtained his PhD in Statistics and Applied Mathematics from INRIA and Grenoble University in 2016 with a thesis on dimensionality reduction. From February 2015 until October 2016, he was visiting researcher at the department of Statistics, TU Graz. Since February 2017, he is a Data Scientist at Know-Center, Graz, Austria.



## 2.8 Machine Learning and Automated Decision Making

Partners of Work Package 3 of Semi40

Author: Anja Zernig; KAI GmbH

Author: Ashwini Pandeshwar, Roman Kern and Manuela Rauch; Know-Center GmbH

*Prologue was initiated by Dr. Anja Zernig, a data scientist at KAI GmbH, which is a research centre owned by Infineon Austria AG. As a result of this combination, any research activity at KAI is closely related to real world problems coming from the semiconductor industry. Within Semi40, KAI and Know-Center worked collaboratively on automated decision-making concepts, using machine learning and artificial intelligence techniques for the ECS community.*

**In the Semi40 project there are a lot of different concepts being applied, can you please briefly explain the main terms: data science, machine learning and decision making systems?**

In one sentence, data science, where one part of it is using machine learning techniques, is an enabler for more automated decision making systems or processes. While nowadays most decision making systems are semi-automated, i.e., a combination of data- and human-driven concepts, the future goes towards fully automated decision making systems. While single tasks are already automated, a global concept is not yet implemented for reasons like legal responsibilities which have not yet been clarified – key word: autonomous driving.

**How can these technologies be applied to help the semiconductor industry?**

Decision support systems, as the name already reveals, supports the expert in making decisions. Beside others, this is a question of time often invested for repetitive tasks which can be automated.

**How important is data when applying Data science and machine learning algorithms? How does the semiconductor domain differ from other domains with respect to the data sets?**

The semiconductor industry is one of the most high-tech industries, consisting of high-tech equipment. Hence, also the possibility to gather ever more data is given and already common practise. Nevertheless, only collecting the data which is recorded is not enough. Concepts of data handling, including a flexible but still structured system on how different data and data types are stored and the subsequent accessibility to this data to enable any automatization are needed.

**Can the role of a domain expert (e.g., process engineer), and the role of a data scientist be done by a single person, or are they working independently from each other?**

Actually, data science combines multiple disciplines: Data analysis, computer science, domain expert knowledge. While the first two disciplines are mostly independent of the application area, when coupled with domain knowledge, it directly provides value for the industry in terms of targeted interpretation of the findings. Therefore, my recommendation is a close cooperation between these disciplines and experts, in order to exchange their specific knowledge. Maximum benefit can only be achieved through joint cooperation, which ultimately pays off.

**Which machine learning techniques are currently being used in practice, i.e., what is the state of practice?**

Machine learning is not a completely new concept. The current hype, however, also arose from the growing computing power that is available today. Many machine learning techniques are already known and established in the scientific world for a long time but only a fraction of them is effectively implemented, although the necessary computing power is largely available. In my

opinion, this mostly comes because of a lack of understanding and thus, trust but this can rapidly change in the future. Just think of computer science – hardly any end-user knows how to build or program a computer, but everyone’s using it. The computer science world has been established a while ago and hence, became common, where nowadays enough trust is here to take it for granted.

### **What are the challenges in implementing machine learning in the semiconductor production process?**

Beside the global acceptance of the methodology, data management is the actual enabler for any machine learning technique. Since data is collected for individual purposes and hence, often also stored in different data silos, first, a consolidation of the data landscape is needed prior to any automated decision making system. This also includes the concepts of data availability and data accessibility. And not to forget the data quality: Since data are the engine for machine learning techniques – the better the data quality, the better the machine learning output.

### **Introduction**

During the course of SemI40, the partners worked together in tandem to address the various problems encountered in the Semiconductor Industry. The whole process of data analysis starts with the data collection and proceeds towards application of a multitude of approaches to address the issues. In the following chapters, we have tried to put together a set of best practices for data collection, knowledge transfer, machine learning and data visualization, all gathered from our experiences while working on this project.

### **Chapter 1: Best Practices in Data Collection**

This chapter supports all parties involved in collecting and providing data by giving essential guidelines on how to collect, represent and handle data so it is usable for the subsequent tasks in the work package.

#### *Creating and gathering data*

The more time and thought are put into creating an environment allowing for proper data collection, the easier it is in the future for any analytics team to extract valuable information. Therefore, during the planning of production processes, managing the data that will be created should not be an afterthought.

There are a few general guidelines that should be followed when implementing any system that generates or stores data. The most important ones are:

- Use standardized names, codes and formats, and use them consistently
- If data is stored in multiple tables, use a normalized form
- Document all table and column/label definitions
- Document data files and their content and use descriptive file names for exported files

More information and best practices on the topic can be found in [43].

As the use cases in work package 3 focus on improving or reasoning about existing production processes, in most cases it was required to work with the available data. This meant that data was created in whatever format the involved equipment (e.g. production machine, measurement device or ERP system) already provided. This data is called raw data.

During the early stages of WP3, the project partners agreed to focus on

1. Continuous and ordinal numbers (e.g. measurement data, scaled classifications)
2. Categorical Values (e.g. status names, true/false, type names, machine ids)
3. Text

Continuous variables are anything measured on a quantitative scale that could be any fractional number. An example would be something like size measured in mm. Ordinal data are data that have a fixed, small ( $< 100$ ) number of levels but are ordered. This could be for example quality test results numbered from 1 (good) to 5 (failed). Categorical data is data where there are multiple categories or classes, but they aren't ordered. An example for this would be production machines. There are situations where media data (such as pictures or audio) or unstructured data with unknown format might be the only available representation, but in these cases an adequate way to interpret this data and convert it to one of these categories needs to be found. This kind of preprocessing step is highly dependent on the data and outside the scope of general data handling guidelines.

Raw data often contains errors or gaps and needs to be cleaned up before using it for analysis, which will be described in the following section.

#### *Cleaning up Data*

Data cleaning is probably one of the more time-consuming and important steps in data analytics. Providing correct and properly cleaned data is vital to the success and speed of almost any form of analytical effort. The problem with existing raw data is that it's often "untidy" and messy.

A "tidy" dataset is defined by Hadley Wickham as "easy to manipulate, model and visualize, and [it has] a specific structure: each variable is a column, each observation is a row, and each type of observational unit is a table." (See [228]).

As a guideline for data sharing in the use cases, the following points are the most important properties of data sets that make data optimally useable for machine learning:

- Each observation is contained within a single row
- Each variable is always contained/described in a single column
- No variable in the tidy dataset is stored in both rows and columns
- Time and Date Formats are consistent across the data
- There should be one table for each "kind" of variable
- If there are multiple tables, they should include a column in the table that allows them to be linked
- Row and column names need to be labelled consistently
- Missing data needs to be indicated consistently
- Multiple types of experimental units are not stored in the same table.
- If data is shared as excel spreadsheets, no macros are used and none of the relevant information is encoded using color
- For many data types, the processing steps are well documented and standardized. So, the work of converting the data from raw form to directly analyzable form can be automated with readily available tools

A detailed section on preprocessing of data can be found in Chapter 2.

#### *Merging Data*

In most cases, data will be gathered from multiple heterogeneous sources, which means it will need to be combined to a single, consistent data set.

Generally, merging data means to make sure that all data points use the same vocabulary, categories and units for columns which describe the same physical entity. If data needs to be converted for this, reproducibility is important. The commands or scripts used to convert the data should be part of the documentation.

#### *Documentation*

For documenting an existing data set, so called data book and code books are created. These documents should start with a thorough description of the processes involved in collecting the data. After that, each variable and its units are defined and explained in detail. If the data has been anonymized, summarized or altered in another way that, that should also be mentioned.

**Data Book** - A Data Book is used to describe the received data files from the involved industrial partner. Within the data transfer process, customer fills out the Data Book with the description of the different files and data sources and hands it over with the transferred data files. To gain a better understanding of the data right from the beginning of the use case, Data Book and Code Books should be filled out together with the involved partners. With this, it is expected that potential issues discovered during/after data transfer are discovered and addressed in early stages of the project. These issues (e.g. questions which aroused during the data-transfer, missing files) should also be documented in the Data Book. Data Book are also used to describe the relation between the different data sources. In addition, Code Books are used to describe the different data sources in more detail (i.e.: each data source has its own Codebook).

**Code Book** - The Code Books contains the relevant information about the data source. Within the Code Book each column of the data files belonging to a data source are described. Code Books follow a standard format for describing the data source, apart from the COLUMN\_NUMBER, COLUMN\_NAME and DESCRIPTION (and others) the following columns are important and should be filled in order to achieve a complete description of the data source:

- COLUMN\_TYPE: Type of values stored within column (e.g. *String, Number, Date, Time*)
- COLUMN\_FORMAT: The format of the column type (e.g. for Date and Time column types)
- DATA\_TYPE: Type of data of the column (e.g. Discrete, Continuous, Categorical: nominal, ordinal)
- VALID\_CATEGORIES: If the values are Categorical which values are valid divided by Semicolon (e.g. Gender: Male; Female; Quality: High; Medium; Low)
- MIN\_VALUE and MAX\_VALUE: The minimum and maximum value for each column (if relevant and known)
- NULL\_ABLE: Marker if the value is allowed to be null
- NULL\_MARKER: How are nullable values marked (e.g. NULL, N/A, 0, “”)

The Code Book should also contain further information about the data files:

- Units and their description
- A general glossary for domain specific terms

If multiple files and tables are provided, their relationships must be documented in detail. This means that primary and foreign keys must be explicitly stated, and all other links between concepts/tables/experiments are explained.

To simplify plausibility checks, upper and lower bounds should be defined for numerical variables. For categorical values, all possible categories have to be defined. The meaning of either implicit

(e.g. field is left blank) or explicit declarations (e.g. cells containing markers such as "unknown", "NaN", "NA") of missing values have to be declared. The code book may also describe invariants that can be used to execute further plausibility checks on the data or results, for example when one variable will always have a lower value than another one.

The Data and Code books, with the final data set are passed to the data analyst to begin the analysis stage.

## Chapter 2: Best Practice for Data Science

We identified several areas which are very crucial for Data Science and in this chapter, we lay out guidelines for best practices for these areas.

### *Knowledge sources and knowledge transfer*

The SemI40 partners who contributed to this section had similar experiences while working together in an interdisciplinary UC team. Each team consisted of at least one industrial and one scientific partner, providing the problem definition and the data mining methods, respectively. Figure 1 summarizes the workflow, which reflects the common approach of all UCs. This workflow blueprint can be followed in most data science projects. Additionally, a collection of ideas on how knowledge can be efficiently transferred is mentioned.

In the following, the stages DATA, mainly defined by the business domain experts, METHODS, suggested from the data mining experts and EXPERT KNOWLEDGE, coming from both sides, in Figure 1 are explained in more detail:

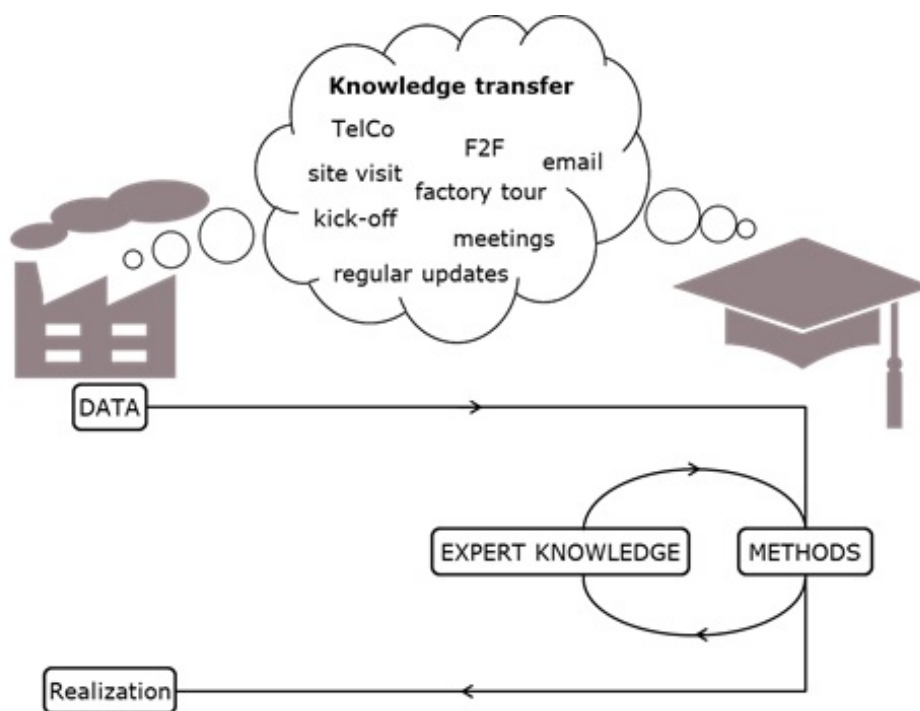


Figure 1: Concept for dealing with different knowledge sources

## Data

- **Data gathering (optional):** If data are not available at the beginning and if there is the possibility to influence the way of how data is collected, partners from the industrial and the academic side should discuss together, which approach to choose for gathering the desired data. In this way, data can be recorded in the needed structure and the minimum sample size for the planned investigations. A good data quality and structure facilitates the analysis by the data scientist. Nevertheless, the collected data should not only be available in a form which is easy to use for a special purpose, but also useful and flexible for further analyses.
- **Data availability:** The availability of data sources must be checked at the very beginning of the project or even before, in the preparation phase. Two scenarios serve as an example here:
  - Data from process steps, handled by external companies, are most probably not available to be analysed.
  - Not at each process step, 100% of the samples are tested but only a subset thereof, and hence, not for each sample complete data are available.

To overcome these limitations, e.g. an own test setup (if possible) can provide the missing data. In addition, simulated data reflecting the reality at the best can provide remedy. One just has to be aware that these artificially generated data might bias the reality and hence it would be better to leave them out or replace them instead. Further, when using simulated data, at the end, an initially proved hypothesis evaluated with simulated data must be extended to a real data set.

- **Data protection:** Data protection is a much-discussed topic. It is not the intension of any industrial partner to provide sensitive information to competitors. To minimize this risk, parameter names can be made unrecognisable and measurement values can be standardized. Nevertheless, the information of these modifications must be provided to the data mining experts in order to avoid misunderstandings and wrong interpretations.
- **Data formats and data sources:** Different data formats or information, available in different sources, further complicates the analysis process because in a first step they must be merged and harmonized. This might sound simple, but needs extra effort. Hence, for stable sources, automated merging would help a lot and is very important for any other subsequent step.

**Methods** - In a first step, when data are handed over from the business domain side to the data mining experts, no expert knowledge regarding the meaning or collection of the data is provided. At this stage, the data mining expert can use completely unsupervised methods and hence, solely let the data speak. This has the advantage that the analyst is not biased in any way and thus, so far unknown correlations between parameters might be detected. The disadvantage, and this is more often the case, is that many spurious correlations are detected, which do not help to solve the problem and lead to unnecessary iterations in the discussion of results. Under these circumstances, motivation suffers. A certain level of detail of information, i.e. expert knowledge to the data, without harming any intellectual properties, would be desirable. Then supervised data mining methods can be used. For supervised methods, knowledge from the business domain expert is integrated by the data mining expert either as prior knowledge or it is used to validate results. Sharing knowledge and experiences with the data mining expert is valuable for an optimal result of automated procedures.

**Expert knowledge** - Expert knowledge is an enabler for good results. Of course, expert knowledge from both sides, i.e. the industrial as well as the academic side, is needed. Good examples are:

- **Knowledge on parameter dependencies** In the provided data sets, most often not only columns

containing measurement values are given, but also calculated columns. Then, dependencies exist between these columns. This is no new information for the data provider and sharing it with the data scientist speeds up the analysis time, because no additional feedback loop is needed. Also, further information, like restrictions of parameters on a practicable range is of interest, since this must be considered in the analysis.

- **Explanation on relevant parameters** Relevant parameters are defined differently from the business domain side or the data mining side. The business domain expert defines those parameters as relevant, which he wants the data mining expert to be analysed. The data mining expert defines the relevance of parameters in the context of statistically significant parameters regarding a special model. In both cases, for the respective parameters, additional background knowledge on these variables is helpful and needs to be explained, at the latest for the final interpretation of results. For instance, the parameter “site number” is a number, indeed, but must not be treated as numerical variable for data analysis. In fact, it acts as categorical information used e.g. for grouping.
- **Required data quality** Dependent on the chosen scientific methodology, appropriate data sets are needed, consisting of, for instance, special data types (cf. clustering or classification) and specific requirements on the sample size (c.f. deep learning procedures). The validity and quality of the analysis or its output largely depends on the quality of the input data, following the slogan “garbage in, garbage out”. To meet the expectations of both sides, one should keep this in mind.

Integrating expert knowledge should always be considered mandatory. If not communicated at the very beginning, it becomes unavoidable for the interpretation of the results. Expert knowledge is needed from both sides since interpretation depends on the business domain, but also on the used data mining algorithms. Close feedback loops and the interpretation of intermediate results helps to steer activities in the right direction and not to proceed with misleading results. High-quality results cannot be expected, if the necessary expert knowledge is not provided. Of course, there must be a balancing between providing too less and too much extra knowledge. This means that no discarding data is provided, which is not useful for the later processing, but also no critical intellectual property is given away, respectively.

**Main success factors** - In this section an excerpt of the lessons learnt from WP3 is given, where the main factors for a successful cooperation and knowledge exchange is stated.

*Planning and coordination* - A joined and detailed definition of the planned tasks is of utmost importance. At the very beginning of the project, the planned tasks have to be divided among the partners to avoid unnecessary overlaps in the methodology. Afterwards, the alignment of organizational tasks, constant project tracking and updates are necessary.

Further, the availability of people involved in the project has to be ensured, because a careful planning of the resources is fundamental in order to assure continuity of the project and to avoid knowledge gaps.

Also, known risks and possible changes in the process have to be clearly communicated to give everyone the chance to rethink their workflow in time. It is important to retain a close and helpful feedback loop, to keep each other on track and to never lose the sight for the bigger picture.

*Communication* - Communication is a crucial factor for progress in any kind of project. On top of formalized meetings and telephone conferences (TelCos), email and unscheduled phone calls are also used to exchange information and clarify topics. This option provides a quick and easy way to communicate without extra scheduling or setting aside time for a meeting. Nevertheless, regular



meetings and at least one preceding Face-to-Face meeting proved to be very valuable within the project communication and the mutual understanding. Further, a common language has to be found since both, the industrial as well as the academic side, have their own terminology.

Sharing and discussing approaches, intermediate results and issues is very important at any stage of the project. This presupposes that the partners are at hand and convey the impression that they are interested in cooperative work.

*Documentation* - Reports containing explanation of techniques and results in an understandable way for both industrial and scientific partners resulted extremely useful to target the final goal and to document progress. It is also important to document tasks which are planned, but also topics which will not be addressed.

#### *The Knowledge Discovery Process*

The success of a data knowledge discovery project is not just dependant on the many algorithms used in the data analysis task. Before attempting to extract useful knowledge from data, it is important to understand the overall approach. This process, known as Knowledge Discovery Process (KDP) defines a sequence of steps (with eventual feedback loops) that should be followed to discover knowledge in data (e.g.: patterns). Cios et al. [40] describe the several reasons to structure a KDP as a standardized process and give an overview of five examples of established Knowledge Discovery Processes. Their analysis classified the process in 3 types of models:

- **Academic Research Models:** Fayyad et al. [54] with a nine-step model developed in 1996 and Anand et al. [6] with an eight-step model introduced in 1998. The model proposed by Fayyad et al. [30] has been used in several different domains including engineering, medicine, production, e-business, and software development. Was also incorporated in a knowledge discovery system called MineSet.
- **Industrial Models:** Industrial models quickly followed academic efforts, two examples are the five-step model by Cabena et al. [191] with support from IBM and the industrial six-step CRISP-DM (CRoss-Industry Standard Process for Data Mining) [34] developed by a large consortium of European Companies [39]. CRISP-DM has become the leading industrial model.
- **Hybrid Models:** Academic and Industrial development needs lead to the introduction of models that combine aspects of both. Cios et al. [40] introduced the six-step KDP model, a hybrid model developed based on the CRISP-DM and adopting it in the academic research.

With a strong industrial support, the CRISP-DM fulfilled all requirements needed in the scope of this project and was used as a guidance for the different phases for the Data Mining and Knowledge Discovery Process in the tasks.



Figure 2: Process diagram showing the relationship between the different phases of CRISP-DM model

The CRISP-DM KDP model (see Figure 2) consists of six steps, which are summarized below:

- **Business understanding:** step focused on the understanding of objectives and requirement from a business perspective. It also converts those objectives and requirements into a Data Mining / Knowledge Discovery problem definition. It can be further broken into sub steps like: determination of business objectives, assessment of the situation, goals, generation of a project plan;
- **Data understanding:** starts with initial data collection and familiarization with data. It also covers the identification of data quality problems, initial insights, interesting subsets, etc. Sub-step examples are: collection of initial data, description of data, exploration of data, verification of data quality;
- **Data preparation:** all activities needed to construct the final dataset (i.e.: the data that will be fed into the algorithms in the next step). It includes selection of data and attributes, data cleaning, construction of new attributes, data transformation, etc;
- **Modelling:** various modelling techniques and algorithms are selected and applied. It involves the use of several methods for the same problem type and calibration of their parameters to optimal values.
- **Evaluation:** : After one or more models have been built and have high quality from a data analysis perspective, the model is evaluated from a business objective perspective. Key sub-steps are: selection of modelling techniques and algorithms, generation of test design, creation of models, assessment of models.

The first 3 CRISP-DM phases will be addressed in more detail in Section 1.5.

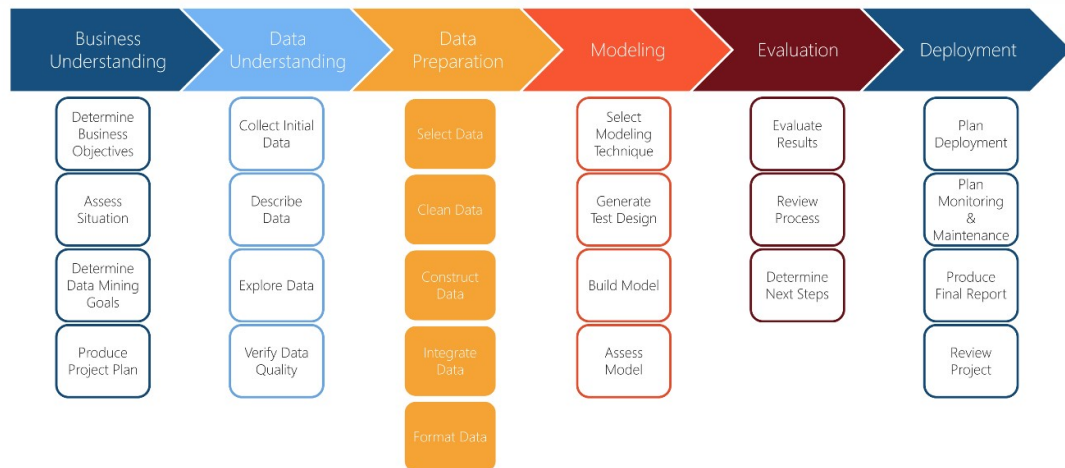


Figure 3: CRISP-DM model phases and tasks (emphasis on the Data Preparation phase)

### *Pre-Processing and Data Cleaning*

Any discipline devoted to the analysis and extraction of knowledge from science engineering requires the collection of data from the field, arriving from different sources and generated by many processes. Typically, this data is collected in the form of symbolic, numerical or textual attributes, obtained via automatic/semi-automatic methods or, still in many cases, via manual processes. Due to this, data often tend to be incomplete, noisy and inconsistent. In other words, we collect dirty data or raw data, that is, if it is not properly pre-processed and cleaned it might pollute all the process of analysis and extraction of knowledge.

In real world data mining projects, it is well known that a significant part of the effort and time is spent on data pre-processing and cleaning. Studies show that this value can surpass 80% of the overall project data mining project effort [192]. In the data acquisition phase, field error rates are typically around 5% or more. Recent studies have also shown that as much as 40% of the collected data is dirty in one way or another [136].

Quality of data plays an important role in information-oriented organizations [217], where the knowledge is extracted from data. Consistency, completeness, accuracy, validity and timeliness are the important characteristics of quality data. Baskar et al. [12] refer to pre-processing as being an important and critical step in the data mining process. Moreover, they refer that success of the data mining task is highly dependent on this preliminary task.

The necessity of data pre-processing is justified by the following needs [12]:

- **Incomplete data:** lacking in attribute values, certain attributes of interest, or data containing only aggregated data. Missing values usually need to be inferred.
- **Noisy data:** data containing errors or outlier values that deviate from the expected. Incorrect data may also result from inconsistencies in naming conventions, data codes, inconsistent formats (e.g.: different date formats). Errors need to be detected and corrected if possible.
- **Inconsistent data:** discrepancies between different data items. Attributes representing a given concept with different names in different data sources. Inconsistency in data needs to be removed.
- **Aggregated data:** aggregated data is often useful, nevertheless in most cases it is not part of the data source, therefore needs to be computed.

- **Enhancing the mining process:** a lot of data does not always mean quality data, often this data is redundant and just slows the mining process. Reducing the amount of data without compromising data quality might speed up the mining process and allows the use of more complex analysis and knowledge discovery and extraction approaches.
- **Improve data quality:** quality decisions must be based on quality data. Detecting anomalies and rectifying them can lead to improve the accuracy and efficiency of the data analysis.

To overcome the previous needs, and with the fundamental purpose of manipulating and transforming raw data so that the information content enfolded in the data set can be exposed, or made more easily accessible, the data pre-processing task should generate a smaller dataset than the original, which allows us to improve the efficiency of the Data Mining process while maintaining the original characteristics of the original data. To perform this task, the following described data pre processing methods are used:

- **Data cleaning:** also known as data cleansing, it attempts to fill in missing values and smooth out noise. It is also responsible for identifying outliers and correct inconsistencies in the data. Incorrect attribute values reasons can be due to faulty or inoperability of data collection instruments, human or computer errors occurred at data entry, transmission errors, and inconsistent naming conventions or data codes. Data cleaning routines should clean the data by filling missing values, smoothing noisy data, identifying and removing outliers, and resolving inconsistency.
- **Data integration:** combines the data from multiple sources into a coherent data set. These sources may include multiple databases and flat files. Major issues with data integration are related with schema integration and entity identification (ex: how the same customer is identified in two different data sources). The use of metadata, that is, data about the data, help in the process of schema integration. Redundancy is another important issue, attributes can be redundant if derived from another data source.
- **Data transformation:** in data transformation, the data is transformed and consolidated into appropriated forms for analysis and processing. Data transformation involves:
  - Normalization: attribute data is scaled to fall within a specific range (i.e.: between -1.0 and 1.0, or between 0.0 and 1.0);
  - Smoothing: remove noise from data using techniques such as binning, clustering and regression;
  - Aggregation: summary or aggregation operations in data for analysis data at multiple granularities;
  - Generalization: low level or 'primitive' (raw) data is replaced by higher level concepts though the use of concept hierarchies (e.g.: age is mapped into young, middle-age, or senior);
- **Data reduction:** complex data analysis and mining on huge amounts of data may take a very long time, making such analysis impractical. Data reduction is either the reduction of volume or reduction of dimensions (number of attributes). The mining on the reduced data set should be more efficient yet produce the same analytical results. Strategies for data reduction include:
  - **Aggregation:** aggregation operations are performed in order to reduce the amount of data. For instance, aggregate a time series in daily, weekly or monthly values;
  - **Dimension reduction:** detection and removal of irrelevant, weakly relevant or redundant attributes or dimensions (examples are t-Distributed Stochastic Neighbour Embedding [133], Latent semantic indexing [44], Principal Component Analysis [97], etc);

- **Data compression:** encoding is used to reduce the data size. Methods used for data compression are Wavelet Transform [141] and Principal Component Analysis [97];
- **Numerosity reduction:** data is replaced or estimated by alternative smaller data representations such as parametric models (store only the model parameters, e.g.: regression and log-linear models), or non-parametric methods (clustering, sampling and histograms);
- **Discretization and concept hierarchy generation:** data values for attributes are replaced by ranges or higher conceptual levels.

### *Data Management*

The management of the data comprises merging data from different heterogeneous and distributed sources, storing and retrieving for data processing. Given the scope, the data will have to step into the following phases in order to be consumed and understood by an algorithm.

Therefore, it is required to collect and combine data from heterogeneous sources and represent it in a format that is adequate for being consumed and understood by an algorithm. This usually also requires cleaning up the data sets and assigning unique labels to all data points. Last but not the least, metadata must be made available by documenting the data set, so the machine learning specialist is able to understand the data while developing an algorithm.

Business understanding - The goal of this phase is to understand what is aimed to be achieved from a business perspective. It includes, among others, the balancing of competing objectives and constraints, and uncover of important factors that could influence the outcome.

As written in the respective task description in the SemI40 ANNEX1 to the Grant Agreement (Description of the Action), domain experts from the industrial partners (and the respective application domains) formalise the process with the emphasis on machine learning tasks that need to be achieved. This formalisation of the process affords the guidance of requirements engineering experts, which have to assure that all use cases can be understood and successfully worked on by the scientific machine learning experts.

There needs to be sufficient transfer of knowledge between manufacturing process experts and scientific data scientists, which is an important success factor for all industrial machine learning projects. To facilitate knowledge sharing between industrial partners and scientific data scientists, an iterative approach has been developed and applied.

Data understanding - The Data Understanding phase requires the data (or access to the data) listed. This initial collection includes data loading, listing the dataset(s) / data sources acquired together with their locations, the methods used to acquire them and any problems encountered. Record problems encountered and any resolutions achieved.

This phase also requires a description of data that has been acquired, including the format of the data, the quantity of data (for example, the number of records and fields in each table), the identities of the fields and any other surface characteristics that have been discovered. Evaluation whether the data acquired satisfies the relevant requirements is also accomplished.

The description of results of this phase, including first findings or initial hypothesis and their impact on the remainder of the project is revealed. A list the results of the data quality verification; if quality problems exist, list possible solutions. Solutions to data quality problems generally depend heavily on both data and business knowledge.

For the specification and documentation, we used a so-called Data Book and Code. Data and Code books have been covered in detail in Chapter 1: Best Practices in Data Collection.

Data preparation - After finishing the Business Understanding and Data Understanding phases, the outcome of these two phases (Data Books and Code Books) is used to construct the final dataset that will be fed into the algorithms in the Modelling phase. This phase called Data Preparation comprise all activities needed to construct the final dataset and includes selection of data and attributes, data cleaning, construction of new attributes, data transformation, etc. The importance of this step also known as Data Cleaning, Data Cleansing, Data Munging, Data Wrangling or Data Tidying was already summarized in Section 1.4.

From Figure 3, the following tasks are part of the Data Preparation phase:

- **Select data:** In this task, the data selection criteria should be reconsidered to decide which dataset will be used. If needed, appropriate additional data (internal or external) should be collected. For the cases where data sources comprise of very large amounts of data, the use of sampling techniques should be considered. It should also be explained why certain data was included or excluded.
- **Clean data:** In this task, techniques that correct, remove or ignore noise should be applied. Decisions on how to deal with special values and their meaning, the aggregation level that data should have (e.g.: time series data aggregated hourly, daily, weekly, etc.) are also taken. The strategy for dealing with missing values should also be devised. Outlier analysis should also be performed at this point.
- **Construct data:** Due to the way specific algorithm expect the input features, it might be needed to generate derived attributes (attribute values that are determined from the values of other attributes). In this task the missing attributes should also be catered for, for example imputed.
- **Integrate data:** Integration of different data sources is a crucial task in the Data Preparation phase. Data is often spread in different locations and formats.
- **Format data:** Final data format might vary due to different needs, very often rearranging attributes (some tools have requirements on the order of the attributes, e.g.: first field being a unique identifier for each record or last field being the outcome field the model is to predict), reordering records (e.g.: modelling tools requires that the records be sorted according to the value of the outcome attribute) and reformatted within-value (purely syntactic changes made to satisfy the requirements of the specific modelling tool, remove illegal characters, uppercase lowercase) are needed.

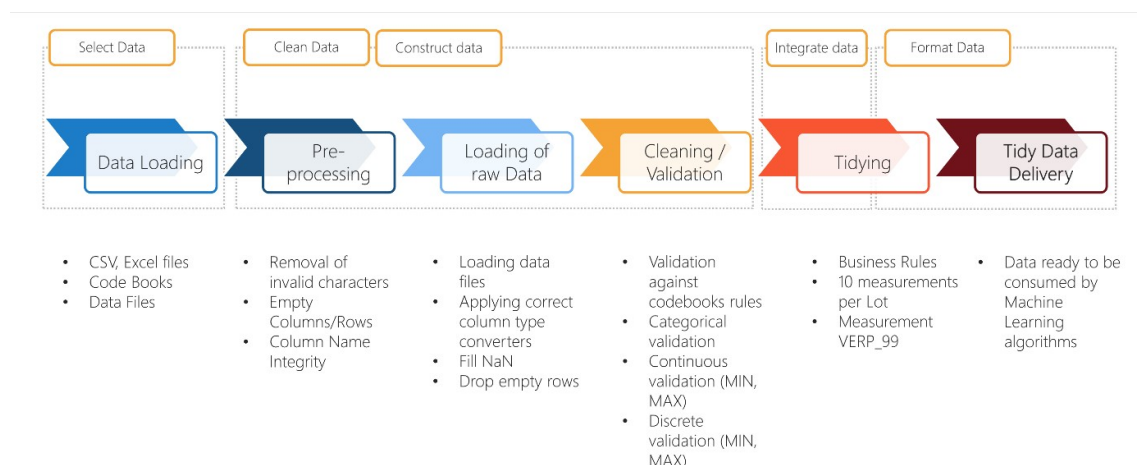


Figure 4: Data Cleaning flow mapping to CRISP-DM Data Preparation phase steps

The diagram Figure 4 shows a tailoring of a Data Cleaning flow to the CRISP-DM Data Preparation phase steps. Please note that tasks are optional and might not occur specifically in one instantiation of a Data Cleaning flow (e.g.: in Figure 3 the task Integrate data has no specific tasks done). Additionally, other tasks can occur in parallel or interleaved, this is the case of the task Clean Data and Construct Data in the example given. A brief description of the concrete tasks done in each step is also presented.

The data flow presented in Figure 4 can be generalized to any data source if correct Data Books and Code Books are provided by the Business Understanding and Data Understanding phases. Ideally this data flow could be automated by loading, interpreting the Data Books and Code Books, deriving rules from them, and performing the Data Preparation phase based on those rules.

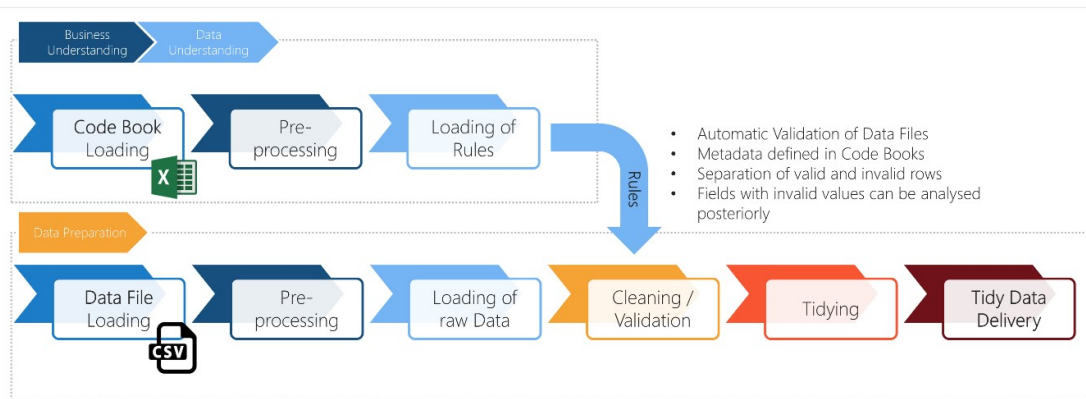


Figure 5: Data Cleaning flow with rules devised from Business and Data Understanding

#### *Joint Architecture for the Processing Pipeline*

Earlier in this chapter, we discovered the need to adopt a Knowledge Discovery Process, the importance of Pre-processing and Data Cleaning, and how a Data Cleaning flow can be tailored into a Knowledge Discovery Process model and fulfil the three primary tasks of the CRISP-DM model: Business Understanding, Data Understanding and Data Preparation. Moreover, an automated way was devised to deliver clean data to algorithms via a conceptual data flow. Business Understanding and Data Understanding is expressed in Data Books and Code Books that will be used in the Data Preparation phase, in an automated way, to clean and validate data coming from one data source. In this section, we are extending the concept of single source Data Cleaning flow to a multiple data source Data Cleaning flow. In addition, the Data Flow will be extended to the Modelling phase of the CRISP-DM model where algorithms will use the clean data from several data sources to perform/support the knowledge discovery process in an integrated processing pipeline. We called this conceptual multi source data flow aiming knowledge discovery from data, **Joint Architecture for the Processing Pipeline**, where techniques from the field of data fusion will be applied, also known as Extract, Transform and Load (ETL).

Figure 6 shows a diagram of the conceptual Joint Architecture for the Processing Pipeline. On the left, each one of the Data Source Cleaning Pipelines is fed with their respective data coming from its source (in this example, a hot folder where files are stored to be processed). First, each Data Source Cleaning Pipeline should be instantiated with the respective Code Book. Only after this step, this specific pipeline will be able to perform data cleaning and validation rules according to the Code Books' specifications. The Data Cleaning Pipeline Detail, shows each one of the steps involved in the processing, cleaning and validation of that specific data source pipeline. Detail was omitted for the other cleaning pipelines, but they have exactly the same processing blocks, but



instantiated with their respective Code Books. In the middle, all Tidy Data coming from all the data sources is delivered to the Join Data Sources step. Here, data is joined into a single data set and prepared to be delivered to the downstream algorithms (Modelling Techniques) that use that data. On the right side are shown four example tasks to be performed in the scope of the SEMI-40 project. The Joint Architecture for the Processing Pipeline supports the following SEMI-40 tasks:

- T1- Correlation Analysis [Pea1895] [148]
- T2 - Reasoning Algorithms (i.e.: supervised ML algorithms such as Support Vector Machines [151])
- T3 - Root Cause Analytics (i.e.: higher-order causality relationships)
- T4 - Advanced ML Algorithms (i.e.: deep neural networks [195])

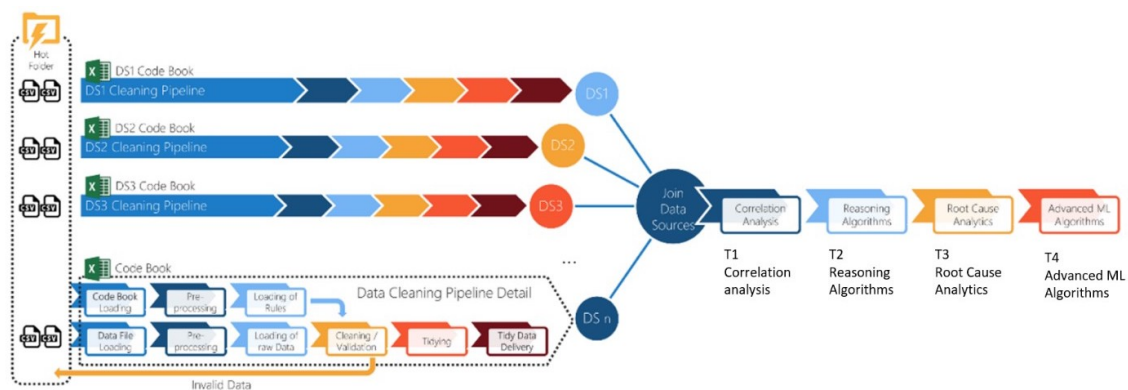


Figure 6: Detail of the Joint Architecture for the Processing Pipeline

This conceptual Joint Architecture for the Processing Pipeline has several properties:

- Receives as inputs Business and Data Understanding rules in the format of Data Books and Code Books
- Agnostic to the input data format. Data of each Data Source should be compliant with their respective Code Book
- Each data source data is clean and validated only according the definitions of the Data Books and Code Books
- Joining of Data Sources is done according to the definitions of the Data Books (i.e.: description of the data sources and their relations)
- Agnostic from the underlying implementation technologies (i.e.: alternative implementation and technologies could be used to implement this Architecture)
- Agnostic and extensible to any Modelling Techniques where the data is used (i.e.: Figure 6 shows four classes of algorithms but the Architecture can be extended to other ones)

To summarize this chapter, pre-processing of data is closely related to knowledge transfer as adequate knowledge of the data is necessary to perform the various steps involved in the preprocessing stage. The guidelines/the best practices listed above have been carefully formulated after observing the various scenarios encountered during the project. An excellent knowledge transfer coupled with great communication and a well processed/tidy data form a perfect raw material for the analytics stage.

### Chapter 3: Best Practice for Machine Learning

Before we jump into the list, we have to mention the first and most important best practice in machine learning – understanding the business problem at hand, determining the expected results (e.g., a classification of items, a forecast of a time series, an estimation of a real-valued number based on features), and establishing the success metrics (e.g., error probabilities or estimation errors). The sheer knowledge of what results you expect out of the algorithms and what you want to do with these results will help you formulate the algorithms and decide what type of data is required for your models, whether the data you have is sufficient, or whether additional data needs to be acquired. For example, labelled data is necessary such that various semi-supervised and supervised algorithms could be applied. In case you encounter data that is not labelled, an expert should be consulted to help label the data. In cases where the labelling is not feasible, only unsupervised methods like clustering, summary statistics, dimensionality reduction (PCA etc.) can be applied.

Data comes in all different sizes, and irrespective of whether it is small data or big data, data pre-processing is a crucial step to address several data quality issues before applying the machine learning algorithms. The importance of data pre-processing has been discussed in detail in the previous chapters. Once the data is pre-processed and ready, the rigorous splitting of the data into train, test and validation sets follows. Extra care must be taken to ensure the split is consciously decided and suits the needs of the algorithms you plan to use. The test set must never be used until the final decisions on the selected methods are made. The training set is used to model the data, the validation set for hyperparameter tuning and evaluating the performance of the model for different combinations of these hyperparameters. Finally, you use the test data to compare your model with other models (you cannot do this with the validation set because it was used earlier to select the hyper parameter values). Hence, the test set helps you compare models in an unbiased way. All said, this again boils down to the discussion of how much data you really have. We will talk about this in the next paragraphs.

The data sets involved under the WP3 of SemI40 can be mainly classified as small data. Small data comes with its own advantages, such as, short data preparation stage and easy application. But one of the major disadvantages is that we can't do much of advanced machine learning, like deep learning, on limited data points. In one of the use cases, non-linear regression could not be used, just because the data set was not big enough to run the algorithm and obtain dependable results. Although SVMs were considered as a possible approach in some use cases, they performed worse than the linear models during the test phase, as they required more data samples to generalize. Had there been enough data points, a wide variety of algorithms could be implemented and tested to achieve some interesting results.

Big data, on the other hand, holds a huge advantage when it comes to advanced machine learning algorithms. Domain agnostic approaches would usually need big data to perform well and produce sound results. Although data preparation takes a lot of time, the availability of a huge number of data points helps to experiment, train and apply across different models. It facilitates complex data mining – prediction, pattern recognition and recommendation. The process of splitting the data into the train and test becomes much more reliable when there are enough data points.

In addition, there are a wide variety of machine learning algorithms available for use, there is no one solution or one approach that fits all. There are several factors that need to be considered while choosing the algorithm. In some cases, it is straight forward as the problem statement is very specific and hence requires a unique approach. But in some cases, the problem statements are open ended and require a lot of trial and errors to find the right algorithm. Finding the right algorithm, that fits the dataset, addresses the current problem statement and ensures reliable results is an important step while using machine learning to address problems. In this project, a wide variety of reasoning algorithms, verification methods and tools were used to address the various problems in the project. The Table 1 gives a comprehensive list of these tools and technologies.

<b>Reasoning algorithms</b>	PCA, ICA, t-SNE, Decision trees, kmeans, VAE, kmedoids, hierarchical clustering, SVD, NMF, Decision Tree Regressor, Lasso, ElasticNet, SGD Regressor, Ridge Regression, Support Vector Regression (SVR), AdaBoostRegressor, PCA, boxplot (statistical data distribution), MBC, kmeans, DBSCAN, SVC, SIR, SVM-regression, Ensemble trees, Decision Trees, GPR
<b>Verification methods</b>	Simulation study, NMI, $R^2$ , entropy, silhouette coefficient, F1, comparison of clustering results with provided ground truth, MSE, MAPE
<b>Implementation</b>	R, MATLAB, Python, Spark

Table 1: Tools and Technologies Used, representing the current state-of-practice

To summarize, although small data is easy to process, and easily obtained, big data holds more advantages when it comes to application of advanced machine learning algorithms. That being said, small data coupled with stronger assumptions and/strong inputs from a domain expert could still be considered as a candidate for advanced algorithms. And above all, sufficient care must be taken into choosing the right algorithm to fit the needs of the problem being addressed.

#### Chapter 4: Best Practice for Visual Data Analytics

Monitoring production in the semiconductor industry requires algorithmic calculations and human interpretations of the results which lead to different actions and decisions. Visual representations have been identified as being suitable in improving the decision-making ability or even enable humans in taking decisions in the first place. In order to identify visual representations supporting human-centered decision-making as well as quantifying influencing factor, dashboards for different use cases containing multiple coordinated visualizations have been created [28] and analysed.

Depending on the data source and applied algorithms, data always has a different format and often requires processing before visualising. Additionally, data is often too large for visualising, therefore further processing is required to reduce complexity. Examples for reducing complexity are data aggregation or filtering for relevant features which require investigation [163]. Furthermore, colour encoding is another important factor when it comes to data visualisation. The ability of users to distinguish different colours is very limited, some people even have larger deficiencies in distinguishing colours. The most frequent deficiency is the so-called red-green colour blindness. As it appears quite frequent, 8% of the male and 0.5% of the female population are affected, it has to be considered when selecting a set of colours for visualising [163].

In order to support users in their decision-making process, dashboards need to support different tasks. Visual data analysis can be divided into four main tasks:

- **Getting an overview:** Enable users to receive an overview of the provided data. If the data is too large, aggregation helps in reducing complexity.
- **Identifying anomalies:** Shows users if and how often anomalies occur in the data as well as whether there are accumulations.
- **Discover and monitor trends:** Trends might be identified over time but can also include modifications of single or multiple parameters to see how the change of parameters affects the result.
- **Unveil data distributions and correlations:** Allows users see the distribution of the whole data as well identifying correlations between distinct parameters.

Often dashboards support some of these tasks but not all. The need to support all tasks highly depend on the requirements.

As the request to analyse data using different devices is increasing, dashboards often support mobile phones and tablets. Different screen sizes including comparable small once require different strategies for displaying visualisations. One possible option is to display only one main visualisation enlarged, while all other visualisations are displayed in the lower or left part as so-called micro visualisations. Users can switch the main visualisation as they prefer and can perform restrictions on it. Restrictions on the main visualisation are also applied on the micro visualisations as they are coordinated with each other. This approach avoids scrolling as all visualisations are visible as well as enables users to quickly switch between the visualisations.

As data complexity is increasing and therefore decision-making is getting more difficult, strategies for collaborative data analysis need to be developed to support the decision process. The approach developed in SemI40 allows to transfer interactions of one user to the dashboard of the other user and vice versa. This approach enables users interactively analysing data on a predefined dashboard while discussing problem statements.

In order to investigate whether a dashboard containing multiple coordinated visualisations support users in performing all required analysis tasks, an evaluation needs to be performed. An evaluation is usually divided into two main steps. The first step is to use the dashboard for data analysis without distinct instructions or by performing pre-defined tasks. Comparing different approaches to identify advantages and disadvantage for specific tasks is a common method which allows comparison [29]. The second step consist of a questionnaire including personal comments and ratings. Important measures for evaluation include mental demand, frustration, stress level and difficulty, defined in the NASA task load index [56].

Summarizing, dashboards for visual data analysis have been identified as being suitable for supporting the decision-making process if they (1) contain all relevant information, (2) reduce complexity and provide details on demand and (3) are intuitive and easy to use [27].

### **Acknowledgments**

This project could not have reached this stage without the active involvement of the WP3 partners. Their contributions are highly appreciated and acknowledged. We wholeheartedly thank each and every partner who was actively involved with the WP3 of this project, participated in knowledge sharing and knowledge exchange, who kept their inhibitions aside and contributed to make this project a success. Thank you for always being enthusiastic and optimistic and striving towards innovation and advancements in the field of semiconductors.

## 2.9 Key Elements of Productivity Enhancement

Author: Author: Germar Schneider; Infineon Technologies Dresden GmbH Co. KG

Author: Thomas Wagner; Technische Universität Dresden

Author: Sebastian Rank; Technische Universität Dresden

Author: Frank Schulze; Technische Universität Dresden

*It has become more and more important for semiconductor manufacturers to optimize their fab performance in the face of current and upcoming trends in the professional and private lives of its customers, like the increasing use of smart phones, autonomous driving, production automation, smart cities etc. Those trends represent challenges concerning the complexity of products, increased demands and a need for excellent quality. This paper describes optimization potentials and solutions in the automated transportation systems and material flow of the fabs. Although innovative solutions and simulation models for transport automation have been presented before, no comprehensive elaboration of self-navigating and free-moving robotic transportation systems exists. Another gap in previous literature this paper is going to fill is the design of a holistic transportation system that combines different types of transportation equipment, like overhead hoist systems and conveyors. Using the results of extensive simulation studies an optimal layout for an advanced hybrid automated material handling system was designed. In addition to that, several solutions for the detection of faults in the transportation system were investigated and implemented, like a measuring FOUP for vibrations. The knowledge gained through the works on the issue of transportation performance enhancement has the potential of revolutionizing production in the semiconductor industry and others by increasing production speed and cycle times, preparing the facility for future advancements in the field of automated transportation and by enabling fab management to quickly detect and remove or prevent failures in the transportation system, thus ensuring wafer quality to satisfy even the most demanding customers.*

### Introduction

Over the last years humans have become more and more dependent on the support of advanced technology in their daily lives. Smart phones, intelligent electronic devices, technologies for home automation and renewable energies and assistance systems in cars are our everyday companions. And we are only at the beginning of this revolution today: According to Nvidia co-founder and CEO Jensen Huang, fully autonomous cars will fully hit the road in 2021 [233]; The penetration of smart home technology in the United States is expected to grow by about 25% to 54% total between 2018 and 2023 [193]; 1.65 billion smart phones will be shipped in 2022 [179]; This results in a compound annual growth rate for the global semiconductor market of 7.76% for the period of 2017 – 2024, adding up to US\$831.5 billion by 2024 [65].

These developments force the semiconductor industry to face the challenge of an immense growth in demand. Finding ways for production optimization and efficiency improvement is a crucial task to be accomplished for the manufacturers to maintain a competitive advantage. Additionally, precise security requirements create the need for a zero failure approach even for tremendously complex semiconductor products that go through hundreds of process steps and are handled in even more transport operations. At the same time these challenges must not present a threat to timely delivery and a competitive cost structure.

This paper presents methods to help establish an optimized production system using the power of digitalization. It aims to take measures towards the envisioned “lot size 1”, starting from cutting-edge transport systems using advanced fab simulation methods to virtualization utilities supporting modern and highly complex manufacturing lines.

**Advanced Transportation Technology at Infineon Technologies Dresden GmbH Co. KG**

The 200mm and 300mm semiconductor production lines at Infineon Dresden are capable of manufacturing a highly diverse product mix with high production volumes, including many products with small lot sizes. Currently, over 400 products are being processed on 40 000 square meters of cleanroom space and transported through the facilities on 12 kilometers' worth of transportation equipment.

The heart of a modern semiconductor manufacturing plant is a fully automated transport system. It must ensure that the respective wafers or lots are being processed at the correct production equipment and that the specific delivery goal required by the customer is being met. Monitoring all transport operations and taking countermeasures in case of deviations are the tasks of a manufacturing execution system.

At the same time, various transport systems are available today, out of which the particular semiconductor manufacturer can choose. Each type of transportation component has specific characteristics resulting in different advantages and disadvantages. The goal is to achieve an optimal fit between the characteristics of the installed components and the requirements and conditions in the manufacturing process of the respective production facility.

**Comparison of Transportation Components**

A variety of choices for transportation system components are at the semiconductor manufacturers' disposal. These include continuous conveyors, rail-guided robots, free-moving robots or overhead transportation systems, the latter of which are typically used in today's 300mm fabs. The choice of a particular system for the fab's specific requirements and conditions alone is a big challenge that requires careful planning and consideration. Due to the constantly changing product portfolio and relatively short product life cycles in the semiconductor industry it is especially important to carefully analyze different possible transport solutions and their impact on all production parameters before starting product development and volume ramps in the fab. Various components were investigated and compared with regard to their individual benefits and weaknesses.

The main advantage of an Overhead Hoist Transportation system (OHT), for example, lies in the space-saving constructions that are possible as the system components can be installed above the production equipment – thereby saving expensive clean room space. Additionally, OHT systems allow very fast transportation movements. On the other hand they require production facilities with a very high ceiling, are cost-intensive and provide lower transportation volume capacity compared to conveyor-based systems. A similar alternative is a Soover-based system (SbT) which is characterized by flat shuttles that move on top of a rail system. Therefore, this kind of equipment might be suitable for buildings with lower ceilings. Conveyor-based systems (CbT) are not quite as fast, but able to handle large shipping volumes in a flexible manner while keeping costs at a medium level. A means of transportation on the ground is the rail-guided vehicle system (RGV). However, since it is not as fast, can only handle medium volumes and is only flexible to a certain extent, this solution is more suitable for automatic loading and unloading of production equipment than serving as the overall transportation system of the fab. Automated guided vehicles (AGV) provide significant flexibility advantages, but are unable to exceed medium speed and volume performance. A cheap and very flexible option is the person-guided vehicle (PGV). However, this option represents a possible quality hazard due to human error and is unable to operate quickly with high capacities. A unified automated material handling system (UAMHS) is as of now a very expensive option, but operates quickly and manages higher volumes. Supporting components like stockers and lifts are not very fast, cannot handle high volumes and are not very flexible. Nonetheless they are necessary components as they serve as connections between different components of the transportation system, are able to connect different floor levels and bridge the timely or spatial gap between two process steps.

### Stability Measures for the Transportation System

Simulation models are very helpful for the analysis of the material flow in the future fab. It enables the identification of bottlenecks, resource shortages and other sources of congestion before the implementation of the transportation systems and before they occur during operation. Since specific products for a wide variety of customer needs are available in Infineon's product portfolio today, small lot sizes ordered by the customer appear more often than they have before. This increases the amount of batch processes in a production environment where the product mix is already very high. Therefore it is particularly important to perform extensive simulation runs concerning the product mix and material transport in the fab before any ramp up steps are taken in order to achieve the optimal tool- and transport system layout.

A new opportunity within the scope of digitalization is monitoring and maintaining such a system in real time instead of relying solely on historical data. Slight operating disturbances, like a defective motor or unintended vibration can already cause a significant damage. With the help of virtualization applications the real material flow can constantly be monitored through newly designed dashboards (see Figure 1). In case of any interruptions countermeasures for instant damage elimination can be taken from a control center.

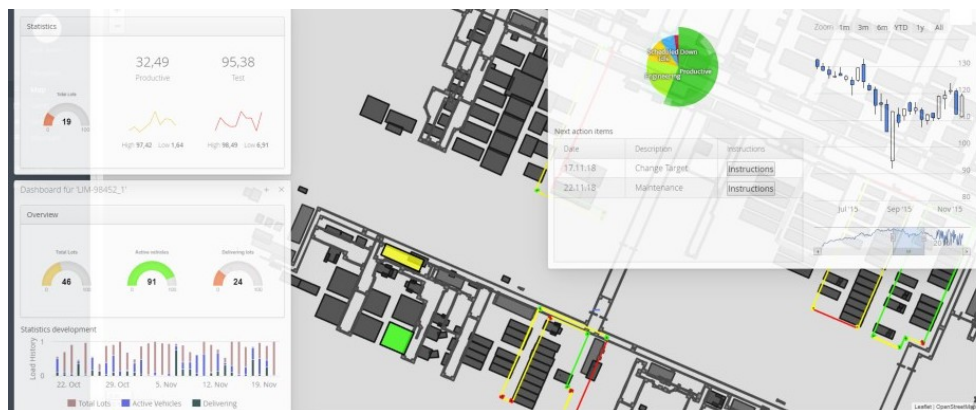


Figure 1: Fab KPI visualisation using dashboards and 2d, map-based approaches

### Results

Figure 2 shows a part of a future fab layout for a bay which has been analyzed through simulation. In the conventional structure of a fab, where production systems with high throughput are grouped within one and the same area, the simulation models show significant congestion. By improving the placement of the equipment, a layout has been achieved where an uninterrupted material flow is possible.



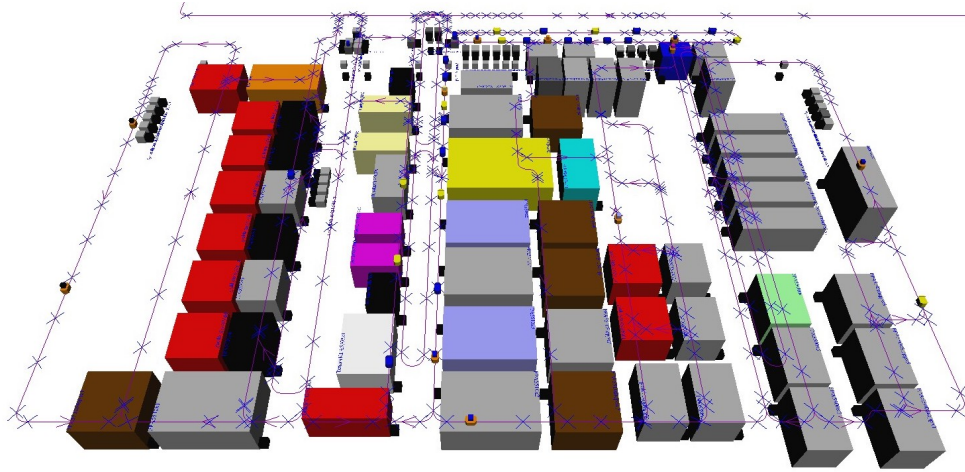


Figure 2: Simulation models with two layout proposals for the evaluation of the tool and transport system layout

At the same time, systems have been developed that allow the discovery of faults, like critical vibration or congestion, in real time. Figure 3 shows a FOUP that has been equipped with a sensor to detect disturbances immediately and to forward them to the control center where a decision can be made regarding an appropriate reaction.

Using the results of the simulation studies combined with specific data processing techniques and visualization of the transportation performance, the optimal design of a powerful automated material handling system was achieved.

## Conclusion

The article described the advancement of the transportation and handling system in a semiconductor fab as the basic requirement for overcoming the challenges resulting from the current trend of digitalization penetrating our personal and professional lives, as well as manufacturing environments themselves.

A carefully thought out transportation system with as little human intervention as possible is able to schedule wafers in the most efficient way, handle them without presenting a quality hazard and generate data at each and every point of the production process. This data can be used to monitor and visualize the current fab performance, detect failures of the system and congested situations before they appear, and to perform root cause analyses and diagnoses in case of quality deviations.

It enables the semiconductor manufacturer to raise his fabs to a level worthy of the term “Industry 4.0” and to maintain a competitive market position through high product quality, high availability and satisfied customers.



Figure 3: Measuring FOUP installation for the recognition of automated material handling anomalies

#### About Authors

Dr. **Germar Schneider** holds a Diploma in chemistry and a PhD in analytical chemistry of high purity materials. He joined the Siemens AG in 1995 in the joint venture of SIEMENS/IBM in Essonnes and worked for three years as a process engineer in the Wet department. In 1998 he changed from Essonnes to Dresden and was responsible as a section manager for the 200 mm Wet department. From 2004 to 2008 he installed a team for factory automation integration, which was a basis for several upcoming factory and automation projects. Between 2008 and 2012 he worked as a manager within the Infineon wafer test department founded in 2008 and was responsible for production equipment engineering. With more than 20 years of experience combining know how of process engineering, production, maintenance and automation he worked as a work package leader within the EU-projects EPT300, EPPL, Semi40, Productive40, and iDev40.

Dip.-Inf. **Thomas Wagner** is a member of the scientific staff at the Chair of Technical Information Systems of the Technische Universität Dresden. He received his M.S. degree in computer science from Technische Universität Dresden. His research interests include modeling, analysis and optimization of logistic systems as well as the automated diagnosis of production equipment.

Dr. **Sebastian Rank** is member of the scientific staff at the Chair of Material Handling at Technische Universität Dresden. He heads the team of Factory Automation. He holds a M. S. degree in Economics and Engineering (Wirtschaftsingenieurwesen) and Ph.D degree in Engineering, both received from Technische Universität Dresden. His research interests include applications of AMHS simulation as well as statistical effects in logistics systems.

Dr. **Frank Schulze** is member of the scientific staff at Chair of Material Handling at Technische Universität Dresden. He heads the team of modeling and simulation. He holds a M.S. and Ph.D degree in Engineering, both received from Technische Universität Dresden. His research interests include the design, and controlling of material flow systems.

## 2.10 Identification of Chiller Models for HVAC Management

Authors: Federica Acerbi; University of Pavia

Authors: Giuseppe De Nicolao; University of Pavia

*One of the goals of the Project Semi40 (Power Semiconductor and Electronics Manufacturing 4.0) is the design of energy-effective strategies for the optimal management of a multiple chillers system in a large Heating, Ventilation and Air-Conditioning (HVAC) plant. Smart energy management is a feature increasingly required by green manufacturing targets. Propaedeutic to the optimal management is the knowledge of a mathematical model of the chiller's energy efficiency. Over the last decade, several energy performance models were proposed and investigated. Nevertheless, model comparisons and validations were often based datasets collected by laboratory tests or provided by chiller manufacturers. The performances predicted by models identified on such ideal data could be overly optimistic with respect to field performances. Hence, in an Industry 4.0 context, real-time information flowing throughout the plant as well as data storage capabilities become key enablers for smart energy management. In particular, processing historical data by means of innovative computational intelligence methods can lead to more accurate prediction of the energy efficiency. To achieve this goal, however, one must ensure the robustness of the estimated efficiency models against the so-called covariate shift phenomenon, i.e. variations in the statistical distribution of the input variables that occur for examples across different months. In fact, differently from well designed laboratory experiments, not only field data may be unevenly distributed across the possible operating points but this distribution may change across months and seasons, making it difficult to obtain a robust model that can be relied upon under all conditions. In this paper, two main issues are considered: (i) the use of extensive field data to validate and compare four models taken from the literature and a newly proposed Gaussian Process model; (ii) a covariate shift experiment: the use of field data to assess the robustness of the five models identified on one-month data when they are used to predict energy efficiency in other months and seasons. The results obtained in this work offer guidelines and insights for the development and calibration of models of water-cooled chillers, to be used in multiple contexts, including energy analysis, improvement evaluation, and FDD strategy development.*

### Introduction

Over the past decades, technological evolution has fostered the development of the so-called smart factory, in which the manufacturing systems monitor physical processes and make smart decisions through real-time communication and cooperation with humans, machines and sensors. In this context, data science techniques play a crucial role in order to extract knowledge from the wealth of field data produced by continuous monitoring and collection of field measurements. Possible tasks include, but are not limited to, optimization of key performance indicator (KPI), predictive maintenance, feedback to design or decision support. The exploitation of these results in the manufacturing and processing facilities, which provide the environment and resources necessary for day to day production, directly impacts productivity as any breakdowns, system failures, or performance problems can result in significant losses. In large semiconductor manufacturing plants, the optimal management of large Heating, Ventilation and Air-Conditioning (HVAC) systems is of primary importance. These systems account for the greatest fraction of total electric energy consumption, amounting to approximately 25-30% of the fab total energy requirement. Between 40% and 60% of this requirement is due to chillers, which are therefore responsible for about 15% of the total energy use. Such energy intensiveness motivates the research on the optimal management of multiple chiller systems, whose prerequisite is the availability of mathematical models of chiller efficiency. Chiller models are typically classified as either semi-empirical (grey-box), e.g. the Gordon-Ng Universal model [69] and its variants, or empirical (black-box). Empirical models,

in spite of their lack of physical insight, have been gaining an ever-wider acceptance due to their prediction accuracy and generalization ability. Among them, one may mention the bi-quadratic regression (BiQ) [231], multivariate polynomial regression (MPR) [89], and multilayer perceptron (MLP) artificial neural networks [212]. So far, the predictive capabilities of alternative models have been assessed and compared on data sets created by laboratory tests or provided by chiller manufacturers as a result of the interpolation of few well-selected data points. In an Industry 4.0 context, the continuous monitoring and collection of field data discloses new opportunities but raises also robustness issues. Indeed, in the machine learning literature, it is known that the statistical distribution of the covariates, i.e. the input variables of the model, plays a key role in determining its predictive capabilities. In industrial environments, characterized by fluctuations of plant operating conditions, the training dataset often covers just a portion of the input space, thus being only partially representative of test data statistics. Under this covariate shift [210], the identification procedure is challenged to find a model that provides satisfactory predictions also when covariates will fall in regions of the input space that are scarcely represented in the training data. Within SemI40, two key tasks were carried out: (i) comparing four literature models (GNU, BiQ, MPR and MLP) and a newly proposed Gaussian Processes (GP) technique [1] on an extensive six-month dataset recorded on a water-cooled chiller which is part of a large HVAC system; (ii) assessing the robustness properties of the five models in the face of monthly covariate shifts occurring in the same dataset.

### The Covariate Shift Experiment

The study used experimental data from a water-cooled chiller which is part of a large HVAC system that supplies cooling water to the Air Handling Unit (AHU) heat exchangers of the Infineon semiconductor fab located in Villach. The dataset is composed by over 4,000 data points collected at different working conditions from June until October 2016. In most models, the input variables are  $T_{ci}$  and  $Q_{ei}$ , the condenser inlet water temperature and the evaporator cooling capacity, respectively. The output variable is the COP (Coefficient Of Performance). Experimental measurements can be represented as dots in the three-dimensional space of coordinates  $(T_{ci}, Q_{ei}, \text{COP})$ . In its simplest form, a mathematical model of chiller efficiency is a function  $\text{COP} = f(T_{ci}, Q_{ei})$  that predicts COP in correspondence of any possible pair  $(T_{ci}, Q_{ei})$ . In graphical form,  $f(T_{ci}, Q_{ei})$  defines a surface in a three-dimensional space. The experimental points are used to calibrate (to train) the free parameters of the model so as to adapt the shape of the surface to the available data. Some test data, not used for training, are used to compare the predictive capabilities of alternative models. In order to assess to what extent the covariate distribution of field data is subject to shifts during the six-month period, in Fig. 1 the monthly datasets are displayed separately (red dots: training data, collected during that month; blue dots: test data, i.e. data collected during the other five months). The visual comparison of the covariate distributions, represented as green surfaces, shows that they are subject to substantial changes across the semester. The most critical month is October, when the input data are concentrated in a relatively small region. The monthly data were used to run an experiment whose goal is to assess the robustness of the considered models under covariate shifts. For a given month, e.g. May, all the five models were trained using only the training data. The identified models were then used to predict the COP of the test data and the RMSE (Root Mean Square Error) of the prediction error computed. The procedure was repeated for all the six months.

### Results and Discussion

Table 1 reports the RMSEs both in training and test for the global six-month dataset and for each month; the last column named 6-fold reports the average of the six monthly RMSEs for each model when identified on one month and then tested on the remaining 5 months. For reasons of space, the three-dimensional plots are reported just for the BiQ (Figure 2) and GP models (Figure 3) that

represent the worst and best performers, respectively. It is interesting to note that the models behave much differently from each other. The GNU model, for instance, is often the worst performer, but it is relatively robust in the face of covariate shifts and its 6-fold RMSE (0.2164) is comparable to its global RMSE (0.1995). The behavior of the BiQ (Figure 2) and MPR models is far less satisfactory: they perform so badly in October that their 6-fold RMSE (0.4415 and 0.3354, respectively) is definitely worse than GNU's one. When, as in October, the covariate shift renders the training data poorly informative, the increased flexibility guaranteed by the additional parameters turns into a drawback, because the extra degrees of freedom are spent to overfit in a small region at the expense of generalization capabilities. The MLP ANN proves acceptably robust: its 6-fold RMSE is 0.1347, the same order of magnitude of its global RMSE, equal to 0.0807. Finally, the GP model (Figure 3) appears rather robust with a 6-fold RMSE (0.0981) only marginally larger than its global RMSE (0.0781). The results are suggestive of a possible beneficial effect of regularization (used in both the MLP and GP models) on the robustness against covariate shifts. As a matter of fact, it is known that regularization matches model flexibility to data informativeness thus helping to prevent overfitting.

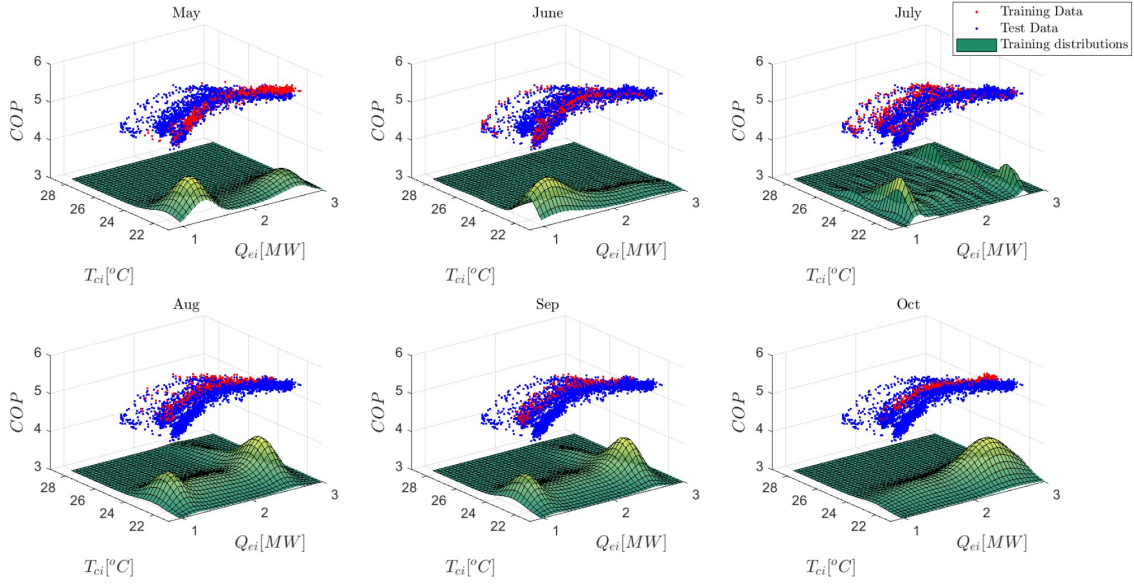


Figure 1: Monthly data sets. Red: training data; Blue: test data; green: kernel density estimates of the training distribution of the input covariates  $Q_{ei}$  and  $T_{ci}$ . Apparently, the covariates distribution undergoes substantial changes across the semester.

	May		June		July		Aug		Sep		Oct		RMSE
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	6-fold.
GNU	0.1640	0.2085	0.1760	0.2782	0.2083	0.1915	0.1851	0.2043	0.1891	0.2031	0.1432	0.2125	<b>0.2164</b>
BiQ	0.0800	0.0942	0.0867	0.0839	0.0720	0.0920	0.0735	0.0881	0.0664	0.0859	0.0612	2.2046	<b>0.4415</b>
MPR	0.0873	0.0992	0.0944	0.1096	0.0836	0.1025	0.0792	0.1237	0.0703	0.1123	0.0670	1.4651	<b>0.3354</b>
MLP	0.0799	0.0931	0.0818	0.0814	0.0705	0.0908	0.0722	0.0835	0.0654	0.0923	0.0599	0.3669	<b>0.1347</b>
GP	0.0792	0.0911	0.0818	0.0823	0.0703	0.0913	0.0720	0.0967	0.0659	0.0941	0.0621	0.1328	<b>0.0981</b>

Table 1: The covariate shift experiment - RMSE on test data



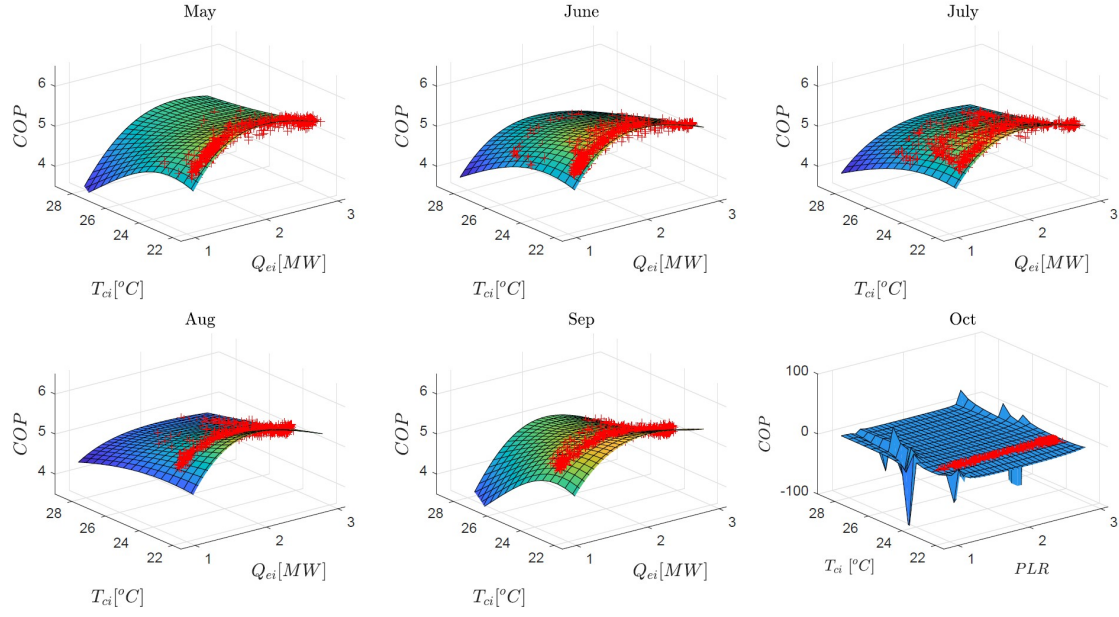


Figure 2: The bi-quadratic model lacks robustness under covariate shifts. In particular, in October its predictive RMSE (2.2046) is about thirty times larger than the RMSE achieved when the model is trained on the entire six-month data set (0.0815). A similar behavior is observed for the MPR model.

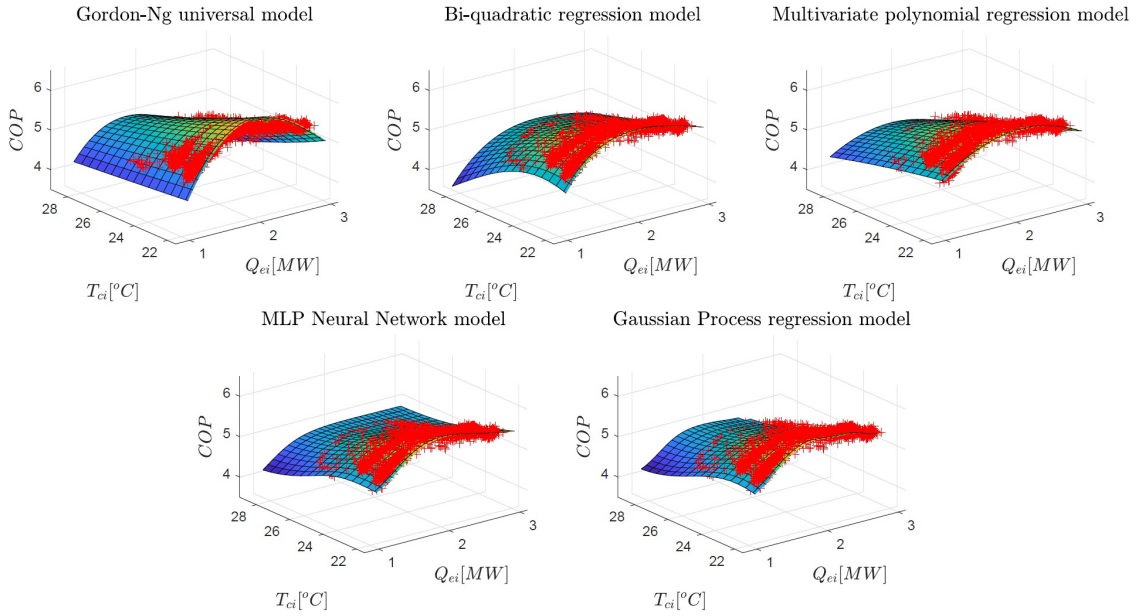


Figure 3: The GP approach not only achieves small prediction errors (see Table 1) but is also robust to covariate shifts. Surface: prediction model; crosses: test data.

### Discussion

The results highlight that caution should be used before abandoning parsimonious and physical-based models, such as the GNU one in favor of more flexible black box models, already proposed in the literature. such as BiQ, MP and MLP models. While in ideal and data rich contexts the

GNU model may well be outperformed, the alternatives may lack the necessary robustness when confronted with real world data whose coverage of all possible operating conditions is far from being optimal. In these cases, the GNU model survived well, while the alternatives yielded poor generalization properties in scarcely sampled regions. On the other hand, machine learning approaches equipped with regularization penalties, appear to be an appealing alternative to the GNU model. Not only they achieve smaller prediction errors but, especially the GP model, prove quite robust to covariate shifts. This suggests that Gaussian processes may represent an innovative and reliable approach to accurately model chiller efficiency in an Industry 4.0 setting, where models are directly obtained and updated from continuously monitored field data.

#### About Authors

**Giuseppe De Nicolao**, born 1962, is Full Professor at the University of Pavia, Italy, Department of Computer and System Science, since 2000. He graduated (cum laude) from the Politecnico di Milano in 1986. His thesis won the annual Quazza Award for the best thesis in Automatic Control defended at the Politecnico di Milano. From 1987 to 1988 he was with the Biomathematics and Biostatistics Unit of the Institute of Pharmacological Researches "Mario Negri", Milano. In 1998 he was appointed a position as a researcher at the Center for System Theory in Milano of the Italian National Research Council. In 1992 he joined the University of Pavia as Associate Professor of Model Identification and Data Analysis. He has published more than 120 journal papers on the following subjects: Control Theory and System Identification mostly in the areas of Filtering, Periodic Systems, Model Predictive Control, Machine and Bayesian Learning, Modelling Identification and Control of Biomedical Systems, Statistical Process Control applied to Semiconductor Manufacturing. Has served the editorial boards of the IEEE Transactions on Automatic Control, IEEE Transactions on Control Systems Technology, and Automatica. He has been the Team Leader of the Pavia Unit in projects funded by state agencies and the European Union as well in industrial research projects funded by multinational companies. He is coinventor of 4 international patents on methods for the analysis of pharmacokinetic/pharmacodynamic data and for the artificial pancreas.

**Federica Acerbi**, got a master degree in Electrical Engineering (Energy Issues branch) from the University of Pavia in April 2016. At present she is a Ph.D. Student in Electronics, Computer Science and Electrical Engineering at the Identification and Control of Dynamic Systems Lab, University of Pavia, Italy. Her research activities are focused on smart production and the development of dynamic simulations which enable production to be planned more exactly and efficiently, as well as quality, capacity utilization and cycle times to be improved. Her main interests are: energy systems, power plants, optimization, machine learning, system identification and modeling.



## 2.11 A Scalable Wafer Dispatching Strategy Based on DP

Author: Alessio Mosca ; University of Pavia

Author: Davide M. Raimondo; University of Pavia

Author: Cristina De Luca; Infineon Technologies AG

*In the semiconductor manufacturing industry, the production process requires wafers to undertake many steps. Suitable dispatching strategies allow to optimally manage the distribution of wafers within the production site, thus reducing delivery time and consequently saving costs. In this paper we develop a new suboptimal real time dispatching strategy for autonomous vehicles responsible for the transport of wafers in a semiconductor production site. Several approaches have been proposed in the literature to solve this problem. Dispatching rules, for example, have been widely used for real-time scheduling because they can provide a very quick and sufficiently good solution. On the other side, choosing the right rules is not an easy task. In this work, a novel approach based on Dynamic Programming (DP) is presented. The proposed solution is applicable to the case where the transport requests are not known a priori, but appear at different time instants. In order to provide a solution which is scalable with the number of orders and the number of robots, we optimize the dispatch of one order at a time. The method relies on the abstraction of the production site on a graph. Differently from standard DP approaches used in this context though, the graph is not static but changes dynamically in order to take into account the dynamic presence of other vehicles on the way. In order to demonstrate the effectiveness of the proposed methodology several simulations have been conducted on a simulator of a semiconductor production site. The latter is implemented using Matlab and is based on a realistic scenario provided by Infineon Villach. Given the current trend in industrial automation, the proposed approach could be of interest not only for the semiconductor industry but for any manufacturing process. Indeed, a scalable dispatching strategy would result in a better exploitation of the available resources (e.g. transport fleet) and consequently reduce the production time/cost required.*

### Introduction

In order to survive in a competitive business environment, semiconductor manufacturing companies are required to meet customer demand exactly in time, quality and quantity. Due to the ever growing demand for electronic products, see e.g. [178], this job has become increasingly difficult over the years. The semiconductor production process is also quite involved since each wafer is required to undertake several steps. For example, a medium-complexity production requires 250-500 steps and uses from 50 to 120 different machines [216]. In order to cope with the complexity of the production process and satisfy customers demand, it is necessary to develop effective and efficient planning, scheduling and dispatching strategies. In the following we focus on viable solutions for the dispatching problem which can be applied in the context of autonomous handling (e.g. mobile robots). In particular, we present a new dispatching strategy based on dynamic programming and graph theory. With respect to previous works (e.g. [145]), we address the case of orders arriving at different time instants in a scalable way. The complexity of a dynamic program approach grows with the number of orders and agents (e.g. vehicles) to be considered. For this reason, we optimize the dispatch of one order at a time and update the graph, abstracting the possible available routes, taking into account the dynamic presence of other vehicles on the way. The proposed method is validated on a scenario reproducing a semiconductor production site taking into account also the presence of charging stations and the need of vehicles to recharge over time.

### Scenario

In the following, we consider a semiconductor production site containing a set of machines ( $\mathbf{M}$ , each of which equipped with a buffer in and a buffer out of respectively size  $bin_m$  and  $bout_m$ ), charging stations ( $\mathbf{C}$ ), Transfer in ( $\mathbf{Ti}$ ) and Transfer out ( $\mathbf{To}$ ) points. Besides the above mentioned points of interest we consider a set of auxiliary nodes  $\mathbf{I}$  used to discretize the routes available to the mobile robots for dispatching operations. Nodes in set  $\mathbf{I}$  are the result of a space discretization performed with a fix discretization step. This latter is chosen in a way consistent with the dimension of the mobile robots and fine enough so to allow collision avoidance and maximize the exploitation of the available space by the fleet of vehicles. Finally, the production site can be abstracted as a directed connected graph  $\mathbf{G} = (\mathbf{N}, \mathbf{E}, \mathbf{w})$ , where  $\mathbf{N} = \mathbf{M} \cup \mathbf{C} \cup \mathbf{Ti} \cup \mathbf{To} \cup \mathbf{I}$  is the set of nodes,  $\mathbf{E} = \mathbf{N} \times \mathbf{N}$  the set of edges and  $\mathbf{w}$  the edge weights representing the travel time between nodes. Given  $\mathbf{G} = (\mathbf{N}, \mathbf{E}, \mathbf{w})$ , the objective is to coordinate a team of robots  $\mathbf{R}$  on the graph so to fulfil the required dispatching in a minimum time. Orders enter the production area from the Transfer in points. From there, they need to be transported to a set of predefined machines for elaboration. When the process is completed, the orders leave the site through the Transfer out points. Since several orders will be managed at the same time, the dispatching strategy needs to select which vehicle to assign to which order taking into account orders priority. Let  $\mathbf{O}(t)$  be the set of orders to be fulfilled at time  $t$  (i.e. orders present at the Transfer in or at the buffer out of one of the machines). Each order is represented by a state a vector containing the following information: number of order, pick-up position and delivery position. Orders must be fulfilled according to their priorities. Top priority is given to orders at buffers out which need to be transported to other machines or to the Transfer out. If several orders of this kind exist, higher priority is given according to the age of the order. Orders at the Transit in come with lower priority. Even in this case, orders are served according to their age. In this work we assume that each robot can transport only one order at a time. Robots have limited battery capacity. Besides, their energy consumption depends on the operation status: in transit, stand-by, recharging. When the battery level goes below a certain threshold, robots are required to recharge at the charging stations ( $\mathbf{C}$ ). Vehicles can be either available (ready to fulfil an order) or not available (busy or recharging/going to recharge). A vehicle can be assigned to an order only if able to satisfy the request and go back to the charging station with the available charge. Note that, in order to minimize the overall time of service, a task can be assigned for a future time to a currently busy vehicle rather than to a currently available one. This can happen when it is faster to fulfil the current order, pick-up the new one rather than having a currently available vehicle fulfilling the new order. At each time instant  $t$ , each robot comes with a vector state containing the following information: Vehicle condition (i.e. in transit, stand-by, recharging), state of charge (SOC), current location, assigned orders, and assigned routes.

### Problem Formulation

In order to find a viable solution to the scenario described above, in the following we consider the case where we optimize only one order at time. Given the order with the highest priority still to be served, this will be assigned to the vehicle which minimizes its delivery time. Let  $o(t)$  be the order with the highest priority at time  $t$ . Define with  $p_0(t)$  and  $d_0(t)$  the pickup and delivery positions of order  $o(t)$ . Let  $R_{av} \subseteq \mathbf{R}$  be the subset of available vehicles at time  $t$ . For each  $r \in R_{av}(t)$ , indicate with  $p_r$ , the position vehicle  $r$  will have when available for fulfilling a new order and  $T_{av,r}(t)$  ( $T_{av,r}(t) \geq t$ ) the corresponding time. Denote with  $T_r(t, p_r, T_{av,r}(t), p_o(t), d_o(t))$  the shortest time required by vehicle  $r$  to reach  $p_0(t)$  and then deliver the order to  $d_0(t)$  starting from  $p_r$  at time  $T_{av,r}(t)$ .

The overall time for service can be obtained as follows

$$T_{r1} = T_r \left( T_{av,r}(t), p_r, T_{av,r}(t), p_r, p_0(t) \right)$$

$$T_{r2} = T_r \left( T_{r1}, p_0(t), T_{r1}, p_0(t), d_0(t) \right)$$

$$T_r \left( t, p_r, T_{av,r}(t), p_0(t), d_0(t) \right) = (T_{av,r}(t) - t) + T_{r1} + T_{r2}$$

$T_{r1}$  and  $T_{r2}$  are obtained using dynamic programming on the graph  $G$ , assuming weights  $w$  indicate the time needed to move between two adjacent nodes.

### Managing Dynamically New Incoming Orders

Finding the optimum for sub-tasks  $T_{r1}$  and  $T_{r2}$  can be addressed using DP. Since the orders come over time and are not known all a priori, it is necessary to find the path on the graph for the new order and its assigned vehicle without modifying the path of the existing ones. While this approach would lead to a suboptimal result, it allows to scale the optimization problem thus providing a solution suitable for real time operation. In order to ensure the obstacle avoidance and consider the fact that some robots are already moving in the space, the optimization needs to operate on a dynamic graph. In particular, nodes occupied dynamically by the previously assigned robots will be omitted by the graph and the optimal dispatching will be performed only on the residual graph. Figure 1 provides an example of this problem where the path of the green vehicle is first optimized. Then the one for the red vehicle is obtained while taking the green one into account.

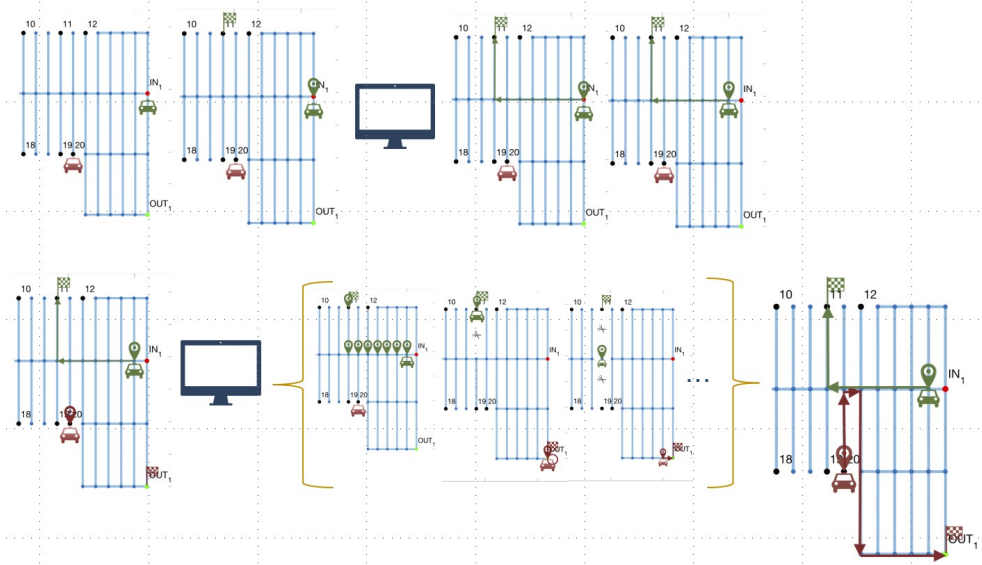


Figure 1: Graphical representation of dynamic programming operating on a dynamic graph

In the DP algorithm for each time step  $k$  the available control action set  $\tilde{U}_k$  is obtained from  $U_k$  by removing all the control action in  $U_k$  that lead to an occupied nodes, only for the time foreseen by the path of the moving robots, thus making an occupied node at a specific given time instant unreachable ensuring that two or more vehicles cannot be in the same node or in the same edge at the same time. Note that, for vehicles which are currently not available, this problem will be addressed starting only from time  $T_{av,r}(t)$  taking into account the vehicles moving on the graph only from that time onwards. Another issue which needs to be taken into account is the future position of the vehicles after completing all the assigned orders. After discussing with Infineon we agreed

on the following: the vehicle remains at the delivery position of the last order and this location is used by subsequent optimizations in order to account for the node occupied by such a vehicle. In case the vehicle in stand-by will be needed for a new order, then this will be involved in a new optimization which will guarantee the collision avoidance with other vehicles without modifying the other pre-defined paths. In case a new order has as delivery point a node occupied by a vehicle in stand-by, then before assigning the new order, this vehicle is sent back to the Transfer In point by solving another DP program. Keep in mind that the state of charge of each vehicle is updated taking into account the length of the paths to be performed and the relative power consumption. Only if the assigned task is achievable within the battery limits this is really assigned to the vehicle. If, after the completion of a task, the residual capacity of a vehicle goes below a predefined threshold, then the vehicle is sent to the charging station. Again, the optimal path to reach such destination is obtained by solving a DP program.

### Results

In order to test the effectiveness of the proposed methodology, we have developed a simulator in Matlab. This latter is based on a realistic scenario suggested by Infineon Austria.

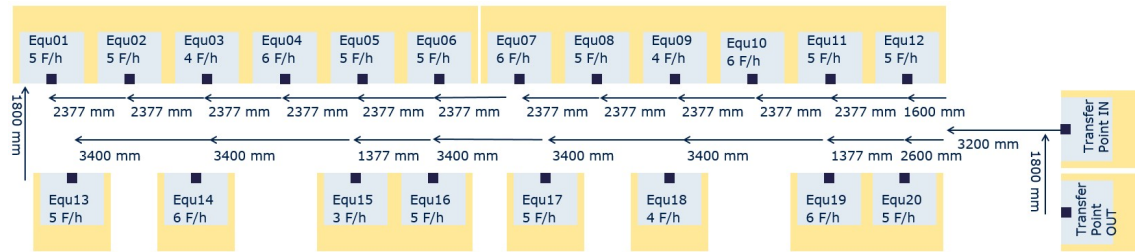


Figure 2: The scenario given by industry

The scenario is composed by 20 machines. The buffer (In-Out) sizes are set equal to the processing capacity of the machines. One Charge station is positioned near the Transfer Point In with the capacity to charge all vehicles at the same time. One Transfer Point in and one Transfer Point out of assumed infinity capacity are considered. A list of orders is generated in accordance with a Poisson distribution ( $\lambda$  [F/h] is the average number of Fous per hour). The machine to which the order is assigned is extracted randomly (uniform distribution). Wafers move towards their destination only when the machine buffer in has space available. Otherwise, they wait in the transfer point in. The elaboration time by a machine is assumed deterministic and equal to  $1/(\text{machine processing capacity})$  (e.g.  $1/(5\text{F/h})=12$  minutes per Foup). After an order is processed, it goes to the machine buffer out. If the buffer out is full, the Foup stays in the machine until the buffer becomes available. The vehicles assignment is made by the dispatching strategy previous proposed. Each vehicle has limited battery capacity and consumes energy unless it is located at the charging station: when the vehicle is moving, it consumes 15% of battery capacity per hour while, if it is in standby only 3%. Robots are required to recharge when the battery goes below 30%. It is assumed that the battery station allows to recharge all vehicles at the same time with a rate of 60% per hour. The graph is obtained in accordance with the maximum robot size increased by 20%. A discretization step larger than the real size of the vehicle was chosen in order to have a safety margin to avoid collisions. The result graph has 1089 (33x33) nodes and 120 edges. The simulation has been conducted using 5 vehicles and the mean time needed for resolving the dispatching problem using our proposed methods is equal to 2.011s which results compatible with the desired sampling time adopted in real scenario. At the following link, <https://youtu.be/kIojrm5dRfU> it is possible to see a short video showing the effectiveness of the proposed approach (the bottom part of the screen indicates the orders currently processed/at the buffer of the different machines).

### Acknowledgments

This work has been performed in the project Power Semiconductor and Electronics Manufacturing 4.0 - (Semi40), under grant agreement No 962466. The project is cofounded by grants from Austria, Germany, Italy, France, Portugal and - Electronic Component Systems for European Leadership Joint Undertaking (ECSEL JU).

### About Authors

**Alessio Mosca** received the Bachelor Degree in Industrial Engineering and the Master Degree (magna cum laude) in Electrical Engineering from the University of Pavia, Italy, in 2013 and 2016, respectively. Currently he is a Ph.D. Student at the University of Pavia, under the supervision of Professor Davide M. Raimondo. His research is mainly focused on optimal vehicle fleet management. From September 2018 to March 2019, he was a visiting scholar at the College of Engineering, Boston University, Boston, Massachusetts, USA, under the supervision of Dr. Calin Belta.

**Cristina De Luca** is graduated in statistical and economical science from University of Padova (Italy). She received a Ph.D. degree in mathematics from the University of Klagenfurt (Austria), 2003. She joined Infineon Technologies Austria AG on January, 2002. She worked on lithography area and transferred the lithography R2R controller from 300mm Dresden to 200mm Villach in cooperation with the University of Klagenfurt. After this she was responsible of different project concerning R2R CMP and CVD controllers in Regensburg (Germany), Kulim (Malaysia) and Villach (Austria) and research for epitaxy R2R controller. She has significant experience in advanced process control, automation and statistical data analysis. She is a certified Project Management since more than ten years. She worked in funding group Villach for more than eight years as project manager of IMPROVE, EPT300, EPPL, and Co-coordinator of SemI40 projects (ENIAC, ECSEL JU). She started a new position in INFINEON Technologies AG, Neubiberg (Germany) on January, 2019 and actually holds the position of Senior Manager Funding Projects Coordination.

**Davide M. Raimondo** received the Ph.D. in Electronics, Computer Science and Electrical Engineering from the University of Pavia, Italy, in 2009. From January 2009 to December 2010 he was a postdoctoral fellow in the Automatic Control Laboratory, ETH Zürich, Switzerland. From December 2010 to May 2015 he was assistant professor at University of Pavia. He has held visiting positions at the Massachusetts Institute of Technology, University of Seville, Vienna University of Technology, University of Konstanz. Prof. Raimondo is currently an associate professor and head of the educational Process Control Laboratory in the Department of Electrical, Computer and Biomedical Engineering at University of Pavia, Italy. He is the author or co-author of more than 75 papers published in refereed journals, edited books, and refereed conference proceedings. He serves as subject editor of the Journal of Optimal Control Applications and Methods and as CEB member of IEEE Control Systems Society. His current research interests include advanced battery management systems, active fault diagnosis and fault-tolerant control, model predictive control and optimization. In 2017, Prof. Raimondo, with co-authors, received the 2014-2016 Automatica Paper Prize Award.

## 2.12 Impact of Test Wafer Flows in Transportation Systems

Authors: David Wittwer; Technische Universität Dresden

Authors: Sebastian Rank; Technische Universität Dresden

Authors: Thorsten Schmidt; Technische Universität Dresden

*The behavior of the transportation system of a semiconductor factory when loaded with a distinct amount of test wafers is fairly unknown. Therefore, the effect of test wafer flows on an overhead hoist transport system was examined in this simulation study with the aim to make a clear statement on the strength and the manner of the impact of test wafer flows. Even though the amount of transports caused by test wafers can be much higher than that of production wafers, surprisingly no research concerning the influence of test wafers on the transportation system was published. A discrete generic simulation model utilizing the simulation tool AutoMod was build based on actual production data with implemented standard dispatching rules to accomplish a highly realistic behavior. The results of this research emphasize the importance of an appropriate consideration of test wafers when analyzing the transport system of a semiconductor factory.*

### Introduction

While for any simulation study researchers aim to model as many features as necessary, they also try to limit them as much as possible. This is especially true for the semiconductor industry and its transportation systems: The systems are the backbones of production and hence directly influence the factory's outcome. So very precise simulation results are needed in order to be able to forecast the systems' behavior and run them efficiently. Considering the modelling efforts and run time performance of the simulation model, however, not every detail of the system can be modelled due to complexity and intricacy. Therefore, the central question in SemI40 is the determination of the level of abstraction of the real system's features and hence the level of model granularity, that has to be applied. In this work special focus is put on test wafers. In the simulation of the transport system of a semiconductor factory a commonly neglected matter are test wafers. They are used in the semiconductor manufacturing process for the qualification of tools, process conditioning and as dummies to fill smaller lots. Regarding test wafers, researchers mainly focus on the reuse and recycle properties of test wafers to avoid buying new wafers or to accomplish stock optimizations (see [165], [58], [90], [230] and Özelkan Cakanyildirim 2006). Furthermore, capacity planning and process automation are examined (see [129] and [53] respectively), but no research concerning the influence of test wafers on the transportation system are published. This is very surprising since especially during ramping or adaption phases the test wafer transports can go up to 80% of the total transport according to our preinvestigations and smaller studies. In 300mm fabs, overhead hoist transport (OHT) systems are frequently used. Therein appropriate vehicles move along rails on the ceiling in order to save expensive space in the clean room environment. OHT systems were subject of examination in a broad range of studies, including this institute's (see [81] and [181]). Due to the wide usage of OHT systems in this paper the impact of test wafers on OHT systems is examined in detail.

### Central Description of the Concept, Activity, Experiments

In order to analyze the effect of test wafers on the OHT system of semiconductor fabs, a fab area with eight fingers and a unified layout served as a test case for a simulation study. Corresponding real-world transaction data of a period of 30 days was collected based on which the transportation routes and process plans were deducted. This data includes information about production wafers, test wafers and development wafers. The latter represent a very small portion of the total amount and are not considered detailed in this study. The transport system of the mentioned area was simulated with an event discrete generic simulation model developed by Rank Hammel (2015)

utilizing the simulation tool AutoMod. The model allows parameter adjustments to enable quick comparisons of alternative configurations. Furthermore, the model includes dispatching strategies similar to the standard Muratec dispatching rules, which results in highly realistic behavior. The chosen amount of runs within the conducted simulation study led to a satisfying level of statistical certainty. The warming up phase has been considered and cut of the simulation results. From a steady state simulation model transportation data of a period of 14 days has been logged/collected and analyzed in detail. Six key figures were defined to measure the performance of the transport system:

1. Amount of test wafers
2. Amount of transports
3. Waiting time of lots before pick-up
4. Throughput
5. Cycle time
6. Vehicle work load

In order to show and quantify the impact of test wafers on wafer fab transportation system three scenarios with a different proportion of test wafers were examined: By doing so different states of the fab (e.g. ramping or high productivity) are taken into account. In the first scenario the full load of all wafers deducted from the transaction data was introduced into the simulation system ("Max"). In scenario two the load of test wafers (and development wafers) was reduced to 50% while the amount of production wafers stayed the same ("50% of Max"). Ultimately the last scenario simulated the OHT and production system with production wafers only ("None").

### Results, Impact

In Table 1 the results for the defined key figures of the three different scenarios are illustrated. Different reactions of each key figure to a specific defined amount of loads of test wafers introduced to the system can be observed. The outcome of the scenarios with test wafer are described as multiples from the scenario without test wafers. Expectedly the amount of transports rises heavily with an increased input of test wafers. The transport time however only changes by a few percent despite the strongly increased load, while the waiting time of processed loads for transport vehicles is increased massively. The increased waiting time however implies high stress on the transportation system and has direct impact on the cycle time of production lots while no correlation between the amount of test wafers and the throughput can be observed. An increased load of test wafers has a quite different effect on the observed key figures. While some key figures only change by relatively small factors, others do change drastically.

Amount of test wafers	None	50% of Max	Max
<b>Transport Amount</b>	1	2.05	2.56
<b>Waiting Time</b>	1	1.38	1.63
<b>Cycle time</b>	1	1.04	1.54
<b>Throughput</b>	1	1.05	0.94
<b>Vehicle Work Load</b>	1	1.71	2.10

Table 1: Key figure results for scenarios with different amounts of test wafer

### Discussion

It is a common practical approach to consider the influence of test wafers simply by a factor derived from the amount of test wafers in the system in order to adjust the capacities. Due to the very different behavior of test wafers compared to production wafers in the system, the mentioned practical strategy is presumed to be insufficient to describe the behavior of the transport system properly. The examined key figures react very differently to the increased load of test wafers



in the transportation system. These different behaviors are difficult to previse and can definitely not be estimated by rough approaches. In future studies the influence of test wafers on the OHT system of a semiconductor factory should be considered to avoid receiving or deducting the wrong conclusions from a simulation study.

#### **Acknowledgments**

A part of the work has been performed in the project Power Semiconductor and Electronics Manufacturing 4.0 (SemI40), under grant agreement No 692466. The project is cofounded by grants from Austria, Germany, Italy, France, Portugal and - Electronic Component Systems for European Leadership Joint Undertaking (ECSEL JU).

#### **About Authors**

**David Wittwer** is member of the scientic staff at the Chair of Material Handling at Technische Universität Dresden. He holds a M. S. degree in Economics and Engineering (Wirtschaftsingenieurwesen) received from Technische Universität Dresden. His research interests include applications of AMHS simulation as well as data based analysis of transportation systems.

**Sebastian Rank** is member of the scientic staff at the Chair of Material Handling at Technische Universität Dresden. He heads the team of Factory Automation. He holds a M. S. degree in Economics and Engineering (Wirtschaftsingenieurwesen) and Ph.D degree in Engineering, both received from Technische Universität Dresden. His research interests include applications of AMHS simulation as well as statistical effects in logistics systems.

**Thorsten Schmidt** heads the Chair of Logistics Engineering at Technische Universität Dresden. He holds degrees in Mechanical Engineering from Technische Universität Dortmund and Industrial Engineering from the Georgia Institute of Technology. His research interests focus on machinery and design of facility logistics and production systems.

## 2.13 Video-Supported Real-Time Analysis of Products and Processes

Authors: Frederic Ringsleben a; University of Applied Sciences Mittweida

Authors: Thomas Haenselmann; University of Applied Sciences Mittweida

*We want to clarify whether and when video processing in industry is useful for production analysis and suggest a novel solution. It is not always possible to make a statement about production errors based on a single image. If errors can be identified during manufacturing, an image may become obsolete. To answer this question, we investigated whether real-time analysis of a manufacturing video is possible under the additional condition that a compact, portable system takes over recording and preprocessing and whether it can be coupled with the company's IT infrastructure. The idea is to analyze a reference video of an error-free production cycle and compare it with subsequent production videos. The analysis refers to the most important information and omits the unimportant information of the video, such as the static background. This approach enables mass inspection of manufacturing processes using small, low-cost smart cameras. In addition, the failure rate of equipment can be reduced if an imminent malfunction is detected.*

### Introduction

Image processing in industry is mostly used for end- or intermediate inspection of products. However, hidden details of finished products cannot be inspected by image processing. This is the case, for example, in the semiconductor industry, where a product is usually manufactured in several steps and consists of several layers. Our approach is not to verify the product but the manufacturing process with a smart camera. A smart camera is an embedded device with integrated camera. The resources of a smart camera are very restricted, so it can only perform small calculations for image processing. The idea is to compare a flawless production cycle with the following production cycles by video. The detection of a deviation gives a hint that manufacturing process may be disturbed. In the field of image processing there are several comparison metrics like structural similarity (SSIM), peak signal-to-noise ratio (PSNR) or mean squared error (MSE) [82] to compare images with each other. The deviation of videos by these metrics is small if they are synchronous in time and the image content is equal. This is the case for video compression. The images of compressed video contain the same content and are not phase shifted. If similar processes are captured, they are not in phase, so the position of a moving object in the video may deviate slightly. However, this can lead to larger deviations of the metrics. Non-changing background also poses a problem because the value of the comparison metric is strongly influenced by large background regions. In the literature, two approaches are usually described to compare videos with each other. One instance is the comparison of single pictures between videos as mentioned above [147]. This is often used to evaluate the information loss caused by video coding. The second approach is a comparison that identifies manipulated illegal video copies on the World Wide Web based only on a few features [157]. The weakness of feature-based comparison for our approach is that only very prominent features are compared, rather than every piece of information in the video. Both comparative approaches are unacceptable to solve our problem. Therefore, we have developed a new approach to deal with the weaknesses of both approaches and in addition to overcome the restriction of smart cameras.

**Central Description of the Concept***Video comparison*

Our central approach is to compare two videos with the temporal brightness variation of their pixels. Therefor we define the term pixel vector as the brightness variation of a pixel in time.

	Compare function of Video pv1 and Video pv2	Advantage	Disadvantage
Euclidian distance	$d(pv1, pv2) = \sqrt{\sum_{t=0}^T (pv1_t - pv2_t)^2}$	easy to compute	Vulnerable to phase-shifted signal
Euclidian distance of discrete Fourier transform	$d(pv1, pv2) = \sqrt{\sum_{t=0}^T (dft(pv1)_t - dft(pv2)_t)^2}$	safe to phase-shifted signal	computing-intensive
Cross-correlation	$d(pv1, pv2)[n] = \sum_{m=-\infty}^{\infty} pv1_m pv2_{(m+n)}$	safe to phase-shifted signal	Vulnerable to signal energy

Table 1: Comparison of different preprocessing approaches

The direct comparison calculates the Euclidean distance between the pixel vectors (see Table 1). The disadvantage of this comparison is the vulnerability to phase shift or jittering of the mutual comparison. The comparison by cross correlation is based on the multiplication of the pixel vectors to be compared, as shown in Table 1. We have noted that the results of cross-correlation depend strongly on the energy of the input. This becomes a disadvantage if the pixel vectors have different energy levels. The comparison in the frequency domain is calculated also with the Euclidean distance after applying the discrete Fourier transform. The comparison in the frequency domain does not have the disadvantage of the previous approaches. It is stable against phase-shift or different energy levels. However, the computational effort of the Fourier transform is greater. Since the disadvantage of direct and cross-correlation comparison exclude their usage, we searched for approaches to reduce the effort needed to perform the comparison in the frequency domain. Therefore, we have tried to cluster similar pixel vectors.

*Video information clustering*

In [172] we present approaches for clustering pixel vectors and discuss problems. The idea is to calculate areas of pixel vectors with similar variance and only compare representative pixel vectors between a flawless production video and a second similar production video. Based on the fact that the two videos are structurally similar, we generate the areas based on the k-means clustering of the flawless video. The primary difficulties of clustering to get coherent areas are the different brightness levels of equally varying pixel vectors and their phase-shift. Background pixel vectors, for example, may vary little but can be very different in brightness and would thus be divided into different clusters. In addition, phase-shifted pixel vectors are also divided into different clusters. We have investigated three pre-processing approaches independently before clustering with k-means in order to obtain coherent areas and overcome these difficulties. These three preprocessing approaches work by means of normalized pixel vectors, by means of the Fourier transform or by the generation of gradients of pixel vectors as described in [172]. After every

preprocessing approach the pixel vectors are clustered by k-means. The k-means clustering returns the average vectors of clusters of pixel vectors called cluster center and the areas of these clusters in the video. Due to the clustering of the cluster centers and the areas, information gets lost. To verify the loss of information of every preprocessing approach, a method is required to measure the loss between the original video and the cluster information. The output of K-means allows us to reconstruct a video and use an established metric like Peak signal-to-noise ratio (PSNR) to measure the information lost. The results of PSNR shows that 128 clusters are sufficient to reach a good value of 30dB with the normalization and frequency approach. This consolidation enables us to keep and process the pixel vectors in the smart camera. In contrast to a video, the 128 pixel vectors fit into the memory of a smart camera and can be processed easily by the camera. A pixel vector of 5 minutes (300 seconds) with a framerate of 60 fps contains 18000 values. Thus, the memory usage of 128 pixel vectors are less than 5Mb. The costs for the Fourier transform and the Euclidean distance calculation are also limited to the 128 pixel vectors and do not pose a problem for a smart camera. However, more powerful computers must do the preprocessing and clustering, because neither the memory nor the computing power of a smart camera is sufficient. Because the effort for preprocessing and clustering only occurs once for the error-free video, we consider this negligible.

### Results, Impact, Conclusion and Discussion

In [171] we have shown that our approach is able to detect various irregularities. We were able to demonstrate that our approach reacts to speed deviations, short-term deviations or irregular vibrations. Table 2 illustrates various deviations of a time-lapse day-night cycle, a Lego Technic scenario and a production process in semiconductor industry from paper [171].






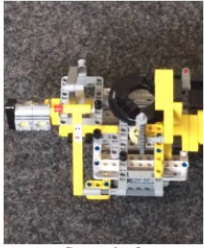
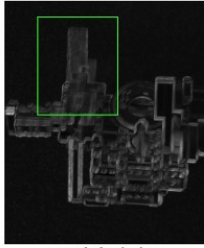
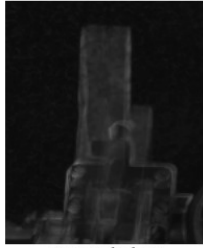


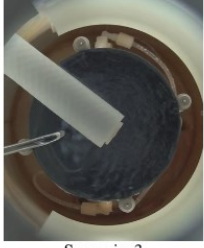
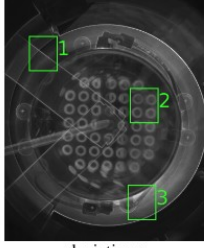

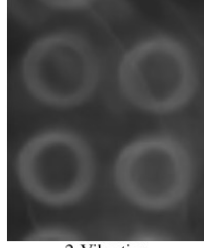
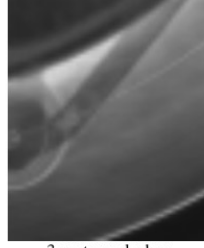
Scenarios	Distance image	Failures		
 Scenario 1	 deviations	 1 various cloudiness	 2 light reflection at the own window	 3 light and reflection at the window
 Scenario 2	 speed deviation	 extremely low	 low	 high
 Scenario 3	 deviations	 1 Edges	 2 Vibration	 3 water splashes

Table 2: Various detectable deviations.

Information is lost by clustering. We determined this loss by comparing the original Video and the clustered values by PSNR. Figure 1 shows the PSNR in dB of the three preprocessing approaches with various numbers of clusters. As a result, about 100 clusters are sufficient to achieve a good value of 30 dB.

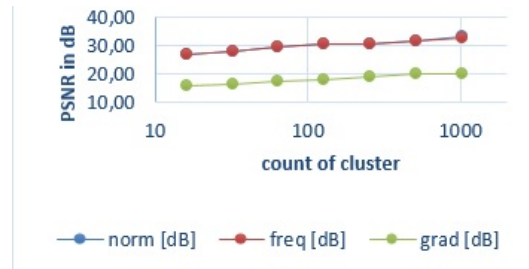


Figure 1: Peak signal-to-noise ratio of video comparison

There are currently few data sets available for this specific case of video comparison. If more special data sets are available, a more detailed investigation can be carried out. This solution was implemented as a demonstrator, but already shows the functionality of this approach.

#### Acknowledgments

A part of the work has been performed in the project Power Semiconductor and Electronics Manufacturing 4.0 - (Semi40), under grant agreement No 962466. The project is cofounded by grants from Austria, Germany, Italy, France, Portugal and - Electronic Component Systems for European Leadership Joint Undertaking (ECSEL JU). Thanks a lot to the sensor team of Infineon Technologies Dresden GmbH for the cooperation within the project.

#### About Authors

**Thomas Haenselmann** is professor at the University of Applied Sciences Mittweida for Computer Graphics and Image Processing. From 1992 to 1998 he studied computer science and business administration at the University of Mannheim. He received his Ph.D. in Computer Science from the University of Mannheim in 2004. He gained work experience at various universities, such as the University of Mannheim, Saarland University (Saarbrücken), Amrita Vishwa Vidyapeetham University, (Kerala, India), National University of Singapore (NUS), NetLab Reykjavik University Iceland, and the University of Mannheim. He is a member of several well-known Technical Program Committees for Multimedia at the ACM or IEEE.

**Frederic Ringsleben** worked since 2016 as researcher in image processing in funding project Semi40. He studied B.Sc. Applied Informatics and M.Sc. Industrial Management from 2010 to 2016. Previously Frederic Ringsleben worked as a mechanic from 2000 to 2010. From 2003 to 2016, he voluntary committed to the German Federal Agency for Technical Relief.

## 2.14 Generic On-Line Video Anomaly Detection for CMP Wafer Cleaning

Authors: Bernat Zaragoza Travieso; Materials Center Leoben Forschung GmbH

Authors: Manfred Mücke; Materials Center Leoben Forschung GmbH

Author: Matthias Fehr; Infineon Technologies Dresden GmbH Co. KG

*Retrofitting existing semiconductor manufacturing equipment with improved monitoring systems is an attractive option to prolong fab lifetime. Machine vision plays a key role in retrofitting since digital image sensors have become ubiquitous and cheap. The core challenges are the need of custom-tailor machine vision applications for each use case and the enormous amount of data. Here, a study is presented, investigating a computationally cheap video similarity metric – ordinal measure – with respect to their applicability in technical process monitoring. Usually, the ordinal measure and related similarity metrics are often used for video copyright monitoring, but have not been studied in the context of technical process monitoring. We use the described method to reduce frames to short feature vectors on an embedded camera. The feature vectors can be used for immediate alarms, post failure analysis and statistical analysis. An implementation of software for on-line video stream segmentation and cutting as well as on-line similarity quantification is shown. Accompanying tools for linking similarity, feature vectors and actual video frames were developed. This software was used to collect and analyse videos from an actual production. The core benefit of the method is that a single low-cost smart camera can be used to monitor existing equipment without the need to install trigger links to current equipment or to custom-tailor the algorithm.*

### Introduction

In many industrial manufacturing processes, there are process stages where the cost of explicit monitoring systems is hard to justify. Variations in these processes typically remain unknown. When performing root cause analysis of detected issues at the end of the line, these process stages are blind spots. The use of non-intrusive, low cost sensors -such as smart cameras- is attractive for retrofitting non-instrumented processing stages. Anomaly detection of industrial processes aims at either assisting operators or automatically assessing the status of a process. In processes where many sensor inputs are available, machine learning algorithms, such as principal components analysis [134], are used to produce plots of condensed data to assist an operator in detecting possible anomalies. In the case of video-only monitored processes, with video surveillance being a relevant case in literature, a common approach is to train models to perform this task autonomously. This is a challenging task with a variety of modelling approaches present in literature, including supervised/unsupervised, normal/abnormal and mixed [223]. Critical aspects of such methods include model training (required amount and type of samples), detection of the cause of a detected anomaly and influence of each model parameter identification. The video comparison method used in this paper uses a single reference video, considered “normal”. An ordinal measure is calculated from every frame and a series of them makes up the video signature [91]. In this work, an exploration to what extent a smart camera in conjunction with a video sequence similarity measure can be used to monitor a wafer cleaning process. The hypothesis presented here is that a fast similarity metric suitable for generic visual anomaly detection and implementable in a cheap smart camera yields a stand-alone smart sensor applicable to many semiconductor process monitoring tasks.

### Central description of the activity and experiments

The process presented here covers from automatic sequence video recording to video analysis. The Sequence Shape Similarity (SSS) algorithm based on Ordinal Measures (OMs) is present in the core of both cases. In video recording, it is used to identify the beginning and end of a wafer cleaning process. In the video analysis, to compare videos across themselves and identify process steps.

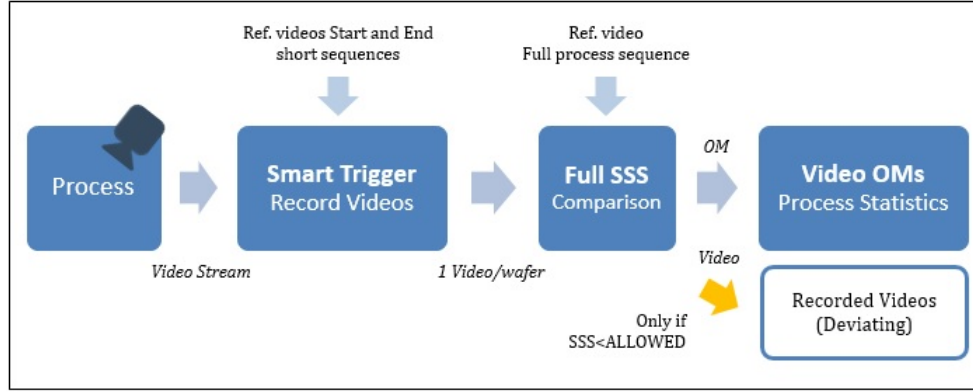


Figure 1: Developed workflow

The analyzed CMP wafer cleaning station is recorded using an iDS uEye BW camera at 30 fps. The camera, hanging from the ceiling on top of the station, has a clear view of the wafer holder and partial views of the device walls, which prevent water from spill (Figure 2). The station cleans wafers sequentially; each cleaning process starts by placing the wafer into the station and ends by removing it. In between, several steps take place, involving wafer spinning, brushing and cleaning with liquid.

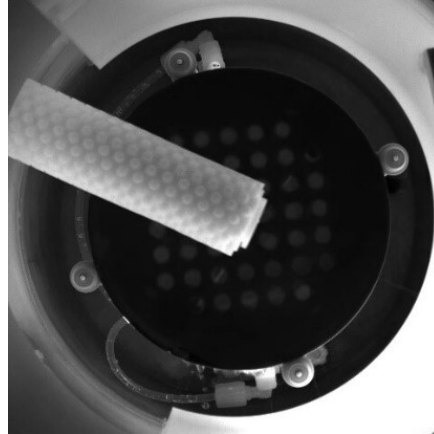


Figure 2: CMP wafer cleaning example frame

For each frame, a feature vector is calculated; a video signature is then a series of feature vectors, one per frame. The OM is a compelling feature due to its robustness [17], which makes it resistant to pixel noise and global light changes. An OM feature vector contains ordered average relative intensities of specific areas of a frame. The selection of the number and shapes of these areas can be used to tailor the OM to a particular process. A regular 3x3 grid is a common generic choice in literature [142] [91] [17]. We extend the common 1D-index ordinal measure to a 2D-index, with the intention to balance the effect of vertical and horizontal changes in a frame. The Sequence



Shape Similarity method (SSS) compares each frame calculating the L1 distance of their OM (1). In the common OM, the cell indexes using a single index number (1D), which results in a higher L1 distance for vertical changes. Specifically, a vertical change influences the L1 distance  $N_x$  times a horizontal one, where  $N_x$  is the number of horizontal blocks (Figure 7). In contrast, with a 2D index, vertical and horizontal changes have the same weight (Figure 8). In the presented algorithm, each frame is divided in a 3 by 3 grid. The intensity values of the pixels contained in each grid cell are averaged and then the OM is calculated. 9 cells produce a 9-length vector OM, where each element specifies the index in which each intensity value is found.

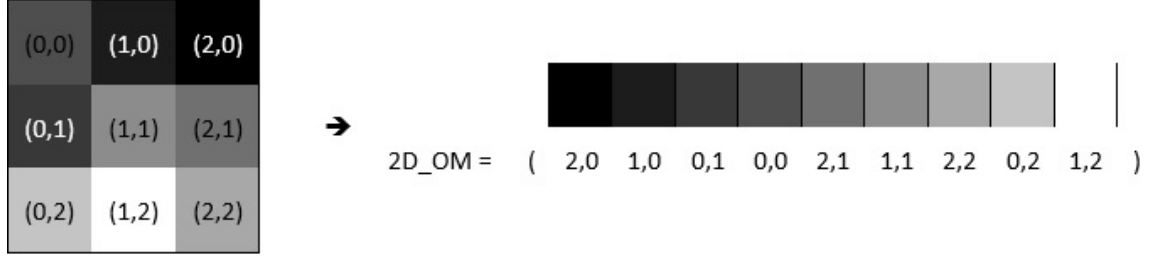


Figure 3: An OM is calculated, for each frame, by ordering the cells index from darker to lighter

In the next step, the Sequence Shape Similarity metric is used to compare two video signatures:

$$SSS(S_x, S_y) = \frac{\sum_{i=1}^M I(Ld(X_i, Y_i) \leq \epsilon)}{M} \quad (2.4)$$

Where  $X_i$  and  $Y_i$  are frames signatures from two different video sequences with the same length  $M$ .  $Ld(X_i, Y_i)$  is the L1 distance between two frames ( $X_i, Y_i$ ) and  $I(x)$  is 1 if  $x$  is true, otherwise 0. Because the same camera is used to record the reference video and the live one, the frames per second are equal and no time resampling is required. Adjusting the value of the threshold  $\epsilon$  allows frames with similar OM to be considered equal while the resulting SSS value  $[0:1]$  denotes how similar the two compared videos are. A  $\epsilon$  of 20% (relative to the maximum distance of 24 for a 3x3 2D OM) has been selected as it successfully ignores differences in frames produced by liquid splatters (Figure 6).

### Smart trigger

To detect a full wafer cleaning process, the video stream is continuously compared with two reference video signatures of the beginning and end of the process. The algorithm uses no other sensor data aside from the video stream provided by the camera. The beginning and end reference video signatures are computed in advance. During execution, the algorithm calculates an OM for every incoming frame and, using the previously calculated  $n$  frames OM, produces a video signature. This video signature length  $n$  needs to be equal to the reference length. The SSS can then be calculated, resulting in a detection if the value is higher than a predefined threshold (Figure 4). The camera stores incoming frames to match the length of the reference beginning sequence. When the beginning sequence is detected (SSS above threshold) in the video stream, the contents of the buffer is stored in a new video file and the following incoming frames are attached to it. The video is recorded until the end sequence is detected. Then, the video file is closed and the algorithm checks for the beginning sequence again. A critical aspect is execution speed. The main loop of the algorithm needs to be executed for every frame, and has to finish before the following one arrives. For the Smart Trigger, this means a maximum execution time of 1/30 seconds. During this time, the following commands need to be executed:

- The incoming frame is stored in the buffer. If the buffer is full, the oldest frame is dropped.

- The incoming frame is also resized to 3x3 pixels (quick grid intensity average), and the OM is calculated and stored in the OMs vector (which length equals to the ref. start or end sequence, whichever is longer). The oldest OM is dropped.
- The SSS is calculated between the OM buffer and the ref. start sequence
  - If  $SSS > 0.35$ , the frame buffer is stored and video starts being recorded
- The SSS is calculated between the OM buffer and the ref. end sequence
  - If  $SSS > 0.35$  and the video is currently being recorded, the recording stops.

### Video set analysis

The video set analysis starts by establishing the SSS threshold (between captured video and reference video) to select how different videos need to differ from reference to be stored. Recorded sequences with an SSS below the threshold will have not only their OM recorded, but also their video. This allows future inspection to aid root cause analysis. If the threshold is set to 1, all videos will be recorded. The computed OM, much smaller than the videos, can be used for process statistics and trends. Aside from the SSS against the reference video, it is possible to calculate the SSS between all pairs of videos using their OM and generate a similarity matrix (Figure 5). This can help to group similar videos together and to identify distant pairs of videos. A more detailed view is obtained by looking at the L1 distance plot between two videos (Figure 6). By looking at the frame number, it is possible to identify in which process steps the bigger deviations have been detected. Finally, we have developed a tool (Figure 7) which allows the user to compare two videos, showing the L1 Distance calculated on the fly according to the selected shift, and even compare them frame by frame. This tool allows to see the results of synchronizing in different times of the video, an interesting feature in processes where some time variation occurs, such as in the wafer cleaning.

### Results

Figure 4 shows the SSS, as used by the smart trigger to detect the placement of the wafer (start of sequence reference) and the removal (end of sequence reference).

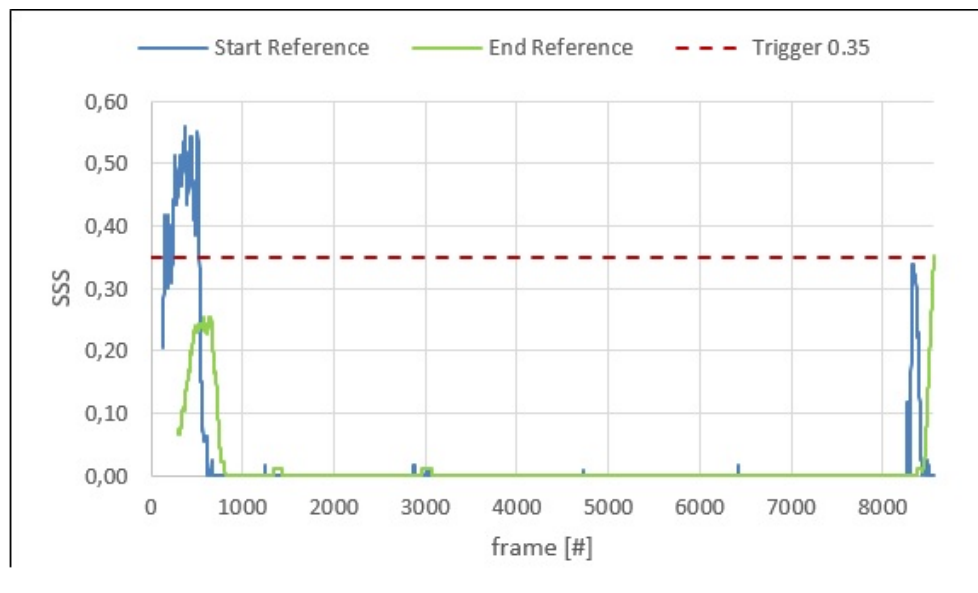


Figure 4: Start and reference SSS across a single wafer cleaning process.

Using the smart trigger, a set of video signatures was collected. The derived video signatures have a much smaller size than the recorded videos (1:1500 for a 480\*480 frame). For example, a video sequences of 4:35min needs about 106 MB (MPEG4) while its video signature is only 72 KB. Due to the small size, video signatures for each wafer could be stored in the process database. Figure 5 shows the similarity (L1 distance) of 80 video signatures ordered in time. Three clusters of videos can be identified.

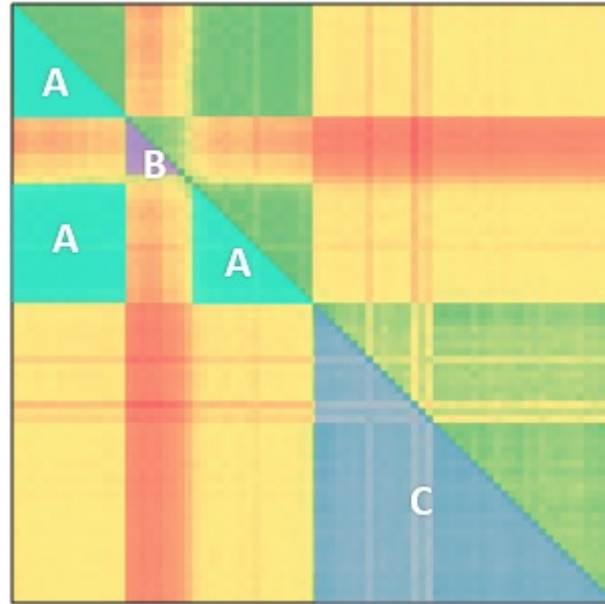


Figure 5: Similarity matrix (Symmetric) of 80 videos. Each intersection is colored according to the calculated SSS between the intersecting videos. Green is 1 (Equal) and Red is 0.61 (most different)

From the similarity matrix, a detailed trend analysis can be performed. Inspection of the per-frame L1 distance can reveal the source of video deviation by process substep. Figure 6 shows an example, revealing the frames 3000..4500 as the main contributors to the SSS.

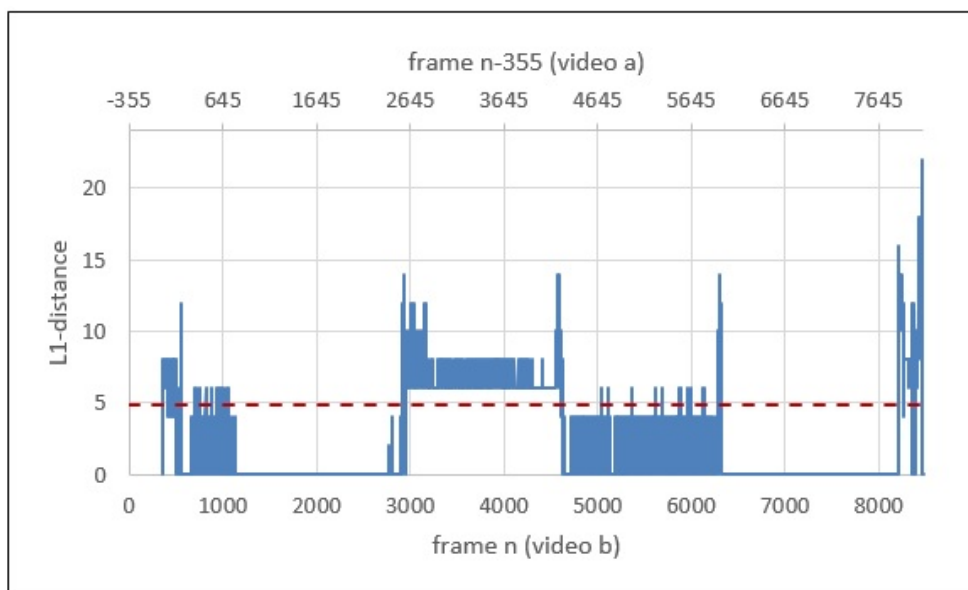


Figure 6: 2D3X3 L1 DISTANCE COMPARISON BETWEEN TWO VIDEOS ( $\epsilon=5$ )

In case the full video was stored for a signature of interest, a detailed comparison with the reference video can be performed. Figure 7 shows a screenshot of the comparison tool with the L1 distance plot on the right and the respective frames of the two videos on the left.

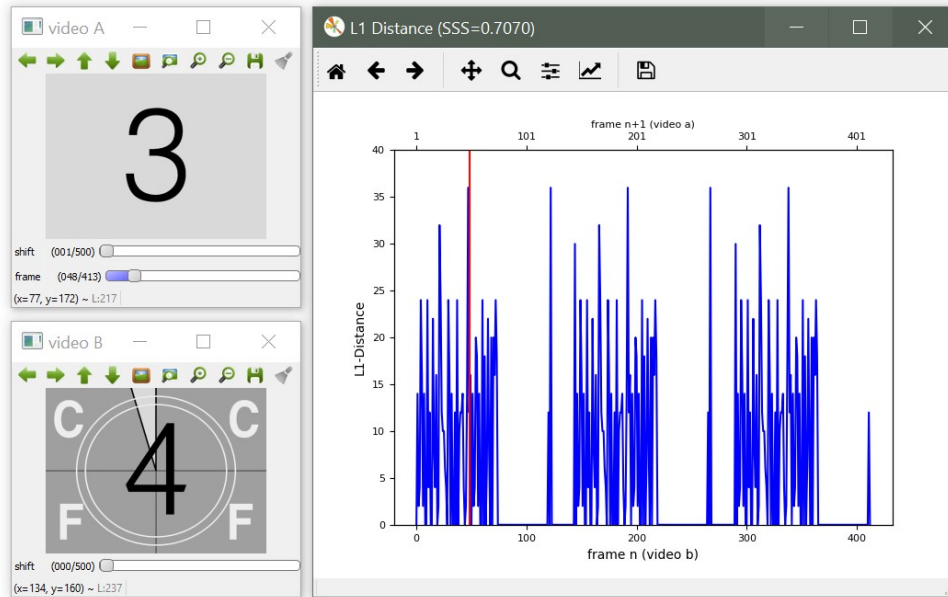


Figure 7: Screenshot of the frame-by-frame comparison tool

### Tables and graphs

0	1
2	3

**A**

0	1
2	3

**B**

0	1
2	3

**C**

$$L1_{frame} = \sum |(0,1,2,3)_A - (1,0,2,3)_B| = 2$$

$$L1_{frame} = \sum |(0,1,2,3)_A - (2,1,0,3)_C| = 4$$

Figure 8: Example of how a vertical change (A-C) has a higher influence than a horizontal one (A-B) in a 1D OM (frame divided in 2x2 blocks)

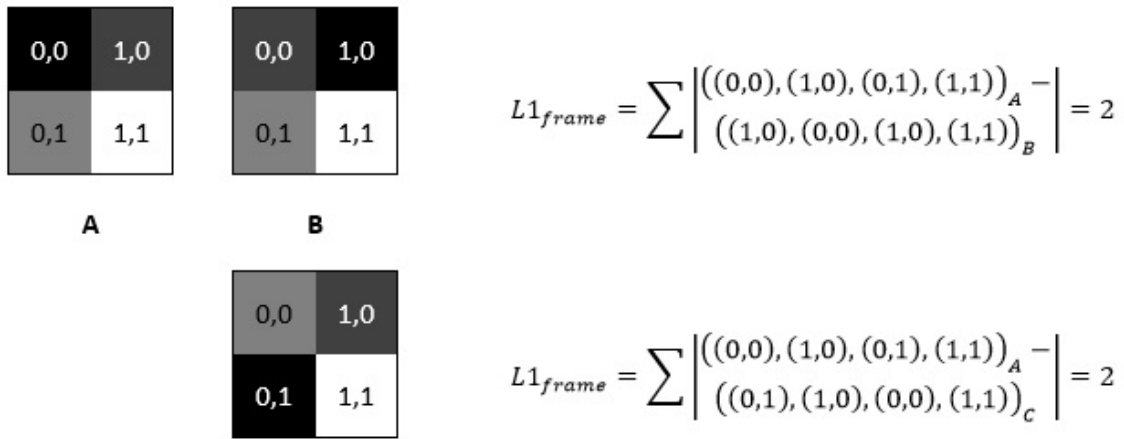


Figure 9: Example of how a vertical change (A-C) and a horizontal one (A-B) have the same weight in a 2D OM (frame divided in 2x2 blocks)

### Discussion

We have shown the use of ordinal measure based similarity metrics for on-line monitoring of a technical process. The implemented Smart Trigger allows for on-line collection of video signatures as well as full videos, if desired. The method allows for cheap instrumentation since no triggers from the existing equipment are needed. OM-based video signatures are small enough to be stored in process databases, enabling trending and guiding root cause analysis. For a new process, only thresholds need to be adapted.

### Acknowledgements

Parts of the work presented have been performed in the project Power Semiconductor and Electronics Manufacturing 4.0 - (Semi40), under grant agreement No 962466. The project is cofounded by grants from Austria, Germany, Italy, France, Portugal and Electronic Component Systems for European Leadership Joint Undertaking (ECSEL JU). The authors also wish to thank Christoph Gaisberger, student employee at Materials Center Leoben Forschung GmbH, for his contribution to the code and plots used in this project.

### About Authors

**Matthias Fehr** works as an equipment engineer at Infineon Technologies Dresden GmbH Co. KG. He is responsible for the improvement of production equipment and the implementation of sensors within the production line for process monitoring. He received his degree as a certified technician, with electrical engineering and data processing focus, from DAA-TECHNIKUM in Jena (Germany) 2004.

**Manfred Mücke** holds the position of key researcher embedded computing at Materials Center Leoben Forschung GmbH, Leoben (Austria). His research interests include inverse models of physical systems, model-based signal processing, machine learning and automatic code generation and their application in next-generation condition monitoring systems. He received his Ph.D. in electronics engineering from Graz University of Technology in 2007.

**Bernat Zaragoza Travieso** joined the Materials Center Leoben Forschung GmbH (Austria) in 2016, where he has worked in enhancing and streamlining gas sensor measurement lab tests, various data science projects and a spectrum of applications ranging from low level code to graphical user interfaces for embedded and PC platforms. He received his MSc in Industrial Engineering, with electrical engineering focus, from the Polytechnic University of Catalonia in 2013 (Spain) and his MSc in Advanced Electronics Engineering from FH Joanneum in 2015 (Austria).

## 2.15 Automated-Generated Long Term Simulation Model

Author: Igor Stogniy; Technische Universität Dresden

Author: Wolfgang Scholl; Infineon Technologies Dresden GmbH Co. KG

*Semiconductor fabs need a simulation model building infrastructure to support data analysis for planning purposes to improve the overall fab performance. In this paper it is briefly described a historical background and a current state of the problem. Several requirements for the simulation model building were formulated. As a core element it was developed the automated-generated Long-Term Simulation model. It is presented the structure of the model and several solutions in data matching and data extraction, which were found during the implementation. The application of the model is a dynamic capacity planning. A work-in-process waves report is shown as an example of the modelling results.*

### Introduction

Today semiconductor fabs have very complex technological and planning processes. The processes have a lot of data which should be analyzed to improve the overall fab performance. One of the ways to support the analysis is to use a simulation. Last twenty years simulation models are widely used in the semiconductor industry for the production planning and control [118]. There is a great need to support decisions in operations, e.g. scheduling of preventive maintenance actions, by a forecast of expected WIP (Work-Process) situation at relevant work centers within the next days. For such applications in operations a highly-detailed, live-data based short-term simulation approach is used. As for the production planning and especially for the capacity planning for a relative long period (up to 6 months) the other approach should be used. It is a long-term simulation. The fact is that the long-term model cannot be an expansion of the short-term model [186]. The long-term model should take into account a data uncertainty which exists always in the future. Ones could not to use detailed data as for the short-term simulation, just because there are not these data for the next months. It means that for the long-term simulation it should be used another more abstract level of the data granularity. And of course the model should be automated-generated, otherwise nobody will use it in a real factory. The data for the long-term simulation are extracted from the fab's data bases and should be converted and uploaded into tables. These tables are the input data for a simulation model. As long as a data volume in a semiconductor fab is huge, it is very difficult and time consuming to extract and prepare the data manually. That is why it was implemented an infrastructure which allows preparing the data tables for the simulation without manual steps. The process of the simulation model building is represented in the Figure 1.

### Historical Background and Current State of the Problem

#### *Simulation*

The history of computer simulation begins with the Manhattan Project, which aimed to develop a nuclear bomb in the US [84]. The next major step towards today's simulation tools was taken at IBM by Geoffrey Gordon in 1960. The first universal simulation system, called the General Purpose Simulation System (GPSS), was developed there [68]. The further development of the simulation is parallel to the progress in computer technology. Several simulation languages and simulation tools had been developed. The choice of the right software solution is closely linked to the actual application and the expected result. Today in the semiconductor industry the simulation is usually divided into four levels (Figure 2).



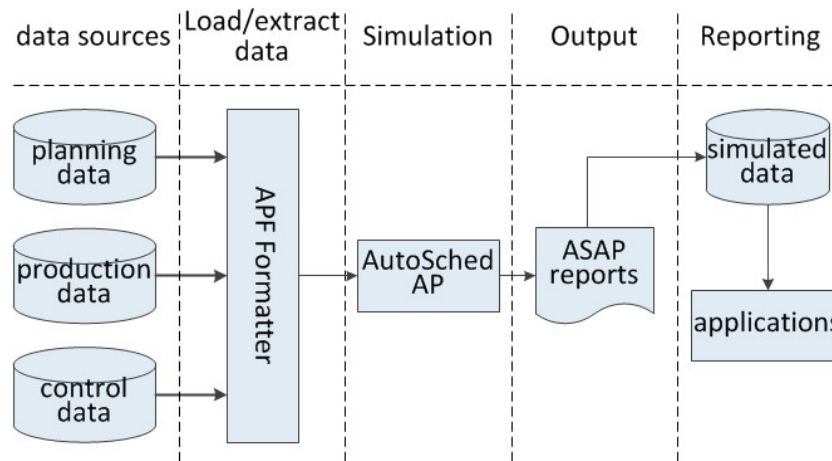


Figure 1: Simulation model building [Stogniy and Scholl 2017]

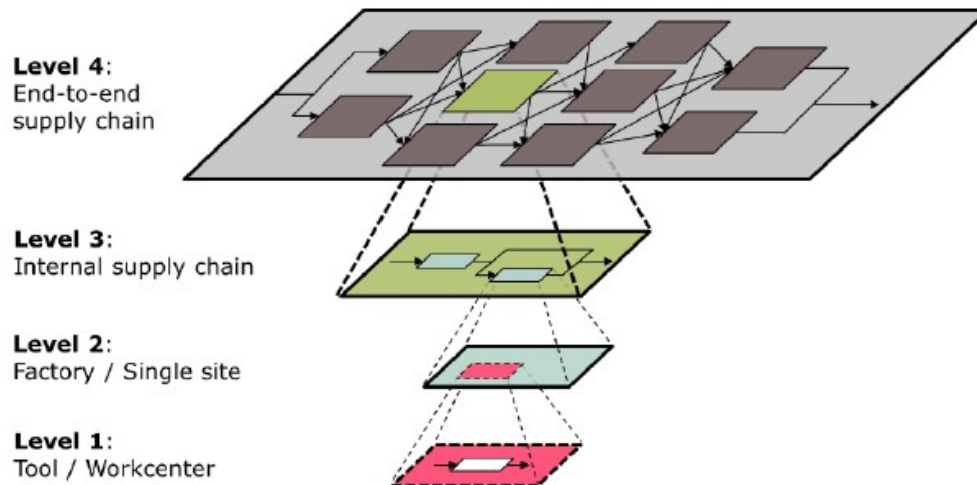


Figure 2: Hierarchy of four simulation levels [Fowler et al. 2015]

The main focus of our research is the level 2 – the fab simulation. The pioneers of factory and facility modelling were larger corporations like IBM or Intel, who soon realized the benefits of simulating. Computer simulation and forecasts helped planning departments make strategic decisions [Kempf 1996]. Currently, for cost reasons, most semiconductor manufacturers use commercial software developed specifically for this type of manufacturing. For the semiconductor industry the problems of model development, maintenance and validation play an important role. There are three different methods in simulation modeling: discrete event, agent based, system dynamics (see Figure 3). In semiconductor industry all three of them are used [177]. To simulate production planning process a common approach is the Discrete Event Simulation (DES). DES is a modelling of event sequences. The occurrence of an event is followed by a change of state. The typical events for a production are, for example, the arrival of a lot at the work center, the end of the process or machine failure. The fact that DES software has proved to be very useful has been confirmed by many leading semiconductor manufacturers (e.g. IBM, Intel or Infineon) in many publications (e.g.: [9], [175], [186]).

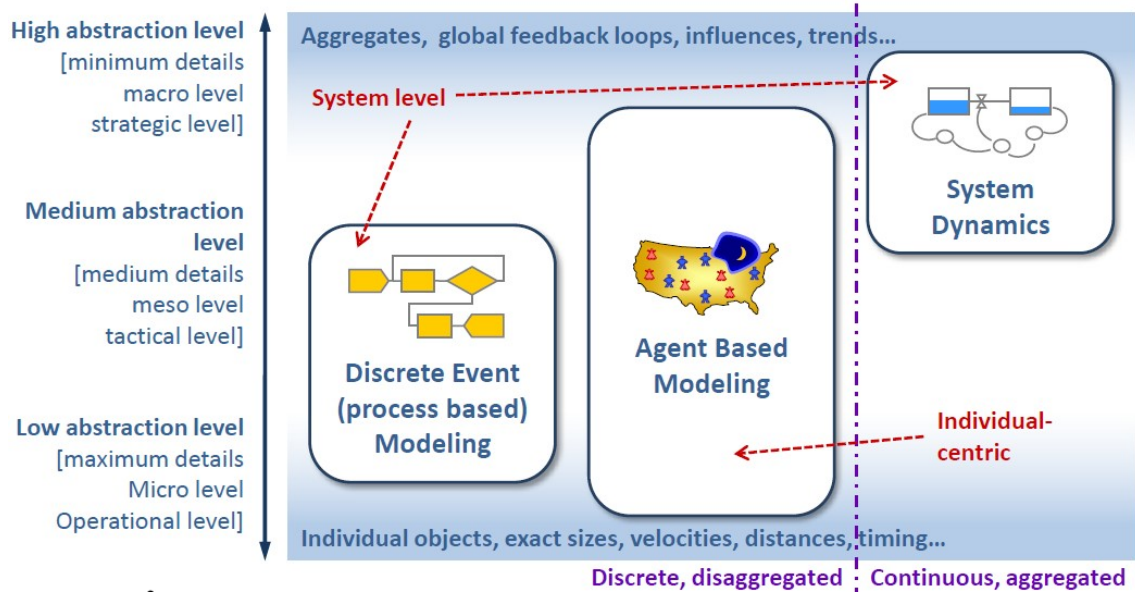


Figure 3: Three methods in simulation modeling [Borshev 2013]

#### *Simulation model building*

Because of the logic complexity and the data volume it is very difficult to build a model manually. It could take up to three days to build a new version of the simulation model manually by a high-qualified specialist. To save time and efforts it is necessary to use an automated model generation in a real fab. This idea is not new. The first approaches were suggested at the end of the 60-s [153]. Later Mathewson defined an automated model generation as software to translate the logic of a model into the code of a simulation model [139]. In the 90-s with the development of computer integrated manufacturing (CIM) system the automated model generation was also developed. It has become possible to build a simulation model based on logic of CIM systems. The simulation model code is generated from a shop floor resource model and a shop floor control model. The shop floor resource model provides much of the static information for the simulation model; while a shop level control model provides much of the dynamic information required by the simulation model [194]. Further the question of automatically generated models rose up [92]. It was investigated an approach to dynamically simulation model building on demand. But this is still an open research question. Automated model generation is still up to date. Today it could be interpreted as a part of Digital Twin, if we mean a twin of a factory, but not of single equipment [73].

#### **Solution description**

##### *Requirements for the simulation model building*

One of the critical requirements for the simulation model building is **maintainability**. In practice, a semiconductor factory is constantly changing: new equipment is installed, old equipment fails down for a long time, and new technological processes are being developed. It has a significant impact on the simulation model building and on the results of the simulation. It is necessary to take into account changes at the factory. The solution is the automated model generation. The **completeness** of the modeling data is the other important aspect of the simulation model building. If the data is incomplete, then the behavior of the model does not match the behavior of the real factory. It increases the likelihood of receiving erroneous model results. Unfortunately in the real factory it is not always possible to get complete data. That is why countermeasures should be taken during the simulation model building. Data **accuracy** is also very important. The accuracy of the model directly depends on the accuracy of the data. If the source data do not reflect reality, then the

simulation results may not be satisfactory (it is well-known principle “garbage in garbage out”). Therefore during the development of the simulation model building master data structure and real data debugging should be at the center of attention. **Data integrity** is one of the quality criteria. Data should not contradict each other. One of the main conditions here is referential integrity. The relationships between different data in different tables must be valid. The analysis of master data plays an important role in the simulation model building. One more important requirement is the **level of abstraction**. It should correspond to the level of detail of the simulation model. However, there is a contradiction: on the one hand, the higher the data specification, the higher is the accuracy of the simulation results theoretically; on the other hand, the higher the data specification, the higher is the complexity and the run time of the model. Therefore, it is necessary to look for a compromise and, perhaps, to use aggregated data. For example, for this purpose it was proposed to use process flow representatives as the routes in the simulation model.

### *Solution*

The data structure for the automated-generated Long-Term Simulation (aLTS) model is based on the real production needs and for the real production planning purposes. It is used planning data (not live data) from a fab data bases. To make the run time of the aLTS model acceptable (1 run less than 1 hour), it was used several simplifications: simplified warm start with only current WIP lot positions; high level of tool configurations and dispatch rules, robust (not flexible) dedications within only important work centers (e.g. lithography, sputtering); using process flow representatives. It was used a stochastic approach (statistics based on historical data) to model sampling and tool downs.

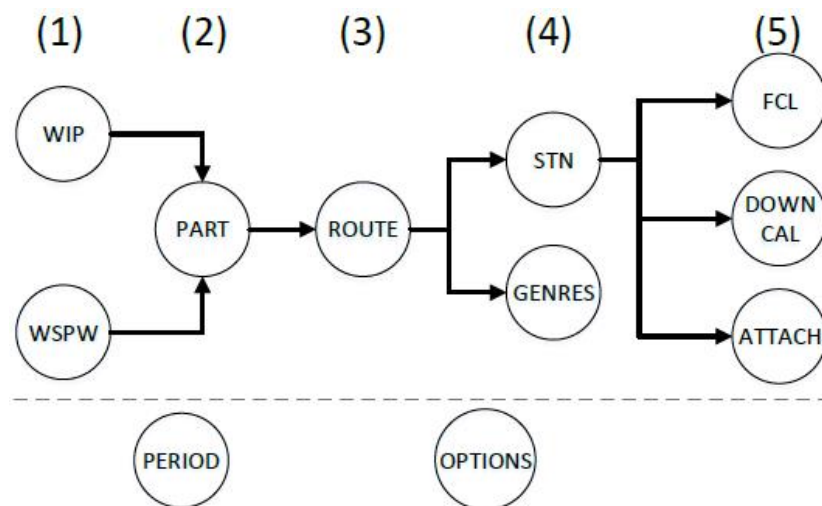


Figure 4: Simulation model files (tables) [Stogniy and Scholl 2017]

In the Figure 4 above five data levels and data flow are shown. Each circle in the Figure 4 is a table of the aLTS. The first level is a demand (orders), which is represented by WIP (Work-In-Process) and WSPW (Wafer Start Per Week). At the second level the product list is described (PART). Level three represents how the products will be produced (ROUTE). At the level 4 there are resources which are used in the steps of the routes: stations (STN) and generic resources (GENRES). The level 5 includes Future Change List (FCL) if there are changes in the components of the model, Down calendar (DOWNCAL) for machine tools (station), Attachment (ATTACH) for connections of different components, like a station with down calendars. Two tables are separate: PERIODs are time intervals at which time-phased parameters can be changed and OPTIONS are different model options. The input data for the aLTS model are planning data and contain uncertainty. That is why it makes sense to use flexible granularity in the model to reduce the influence of uncertainty. The granularity depends on the application and the specific use case. The goal is to reduce complexity of the model and to increase transparency and user friendliness, because of a higher quantity of possible experiments. To develop the aLTS model it was used a well-known simulation tool for semiconductor industry (Autosched AP). In the tool we could find the following chain of entities (core elements):

LOT – PART – ROUTE – STATION FAMILY – STATION

In the granularity model outermost nodes have the same meaning as in reality. LOT is a lot (each lot contain 25 wafers), and STATION is a machine tool (a single equipment). The three nodes in the center are the most interesting. In fact there are a lot of different entities, which are used in Infineon Dresden which could be chosen as the core elements in the simulation. For example as PART ones could choose Process Class, Process Group, Process Line, Product Line, Product Group, Product line, Product Number. Based on expert knowledge Process Line was taken (see Figure 5).

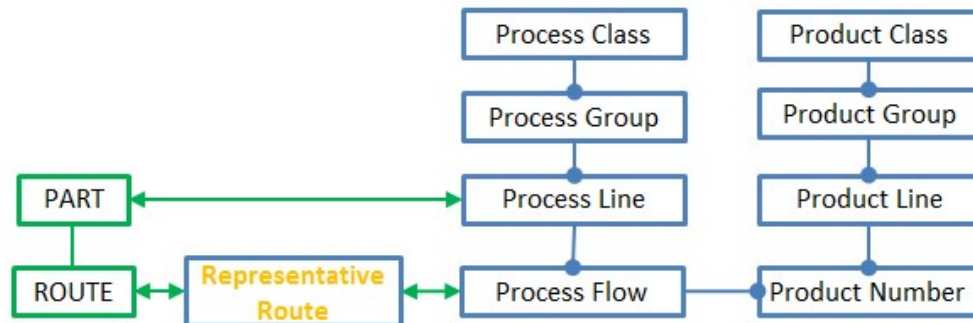


Figure 5: Simulation and real data matching [Stogniy and Scholl 2018]

During the implementation of the data matching a question rose up: how should ones map different core elements if they have a complex relationship? For example in the simulation tool the relationship between PART and ROUTE is 1:1, but in semiconductor fab between Process Line and Process Flow is 1:m. The solution is to make the representative route for each Process Line. It is possible, because there is only small difference between Process Flows for the same Process Line according to the modeling criteria. The most important criterion is duration of route's operations. The other criteria are based on simulation purposes, which are work-in-process, cycle time, flow factor prognosis. The implementation of the simulation model building (see Figure 1) had conflicting objectives. On the one hand it was needed to automate a complicated data extraction process, on the other hand at the same time it should be an easy-use for the non IT-people data extraction. That is why it was used APF Formatter. The tool allows ones to represent SQL-queries in a visualized form. The infrastructure is realized as a set of several visualized SQL-queries in APF Formatter. The overview of the visualized SQL-Queries is represented in the Figure 6.

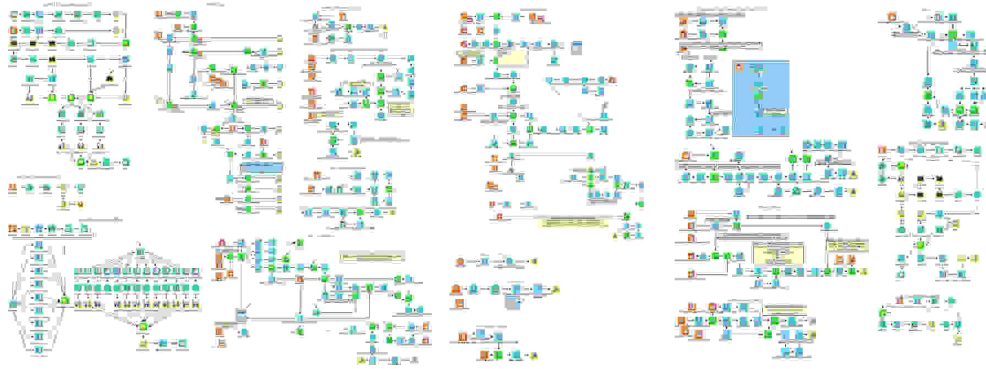


Figure 6: Data extraction and conversion. Visualized SQL-queries.

During the data extraction and conversion several transformation steps were applied: filtering, renaming, and joining. It was necessary to filter out some irrelevant data: dummy values; not up-to-date data; values do not belong to particular facilities; engineering lots etc. Some of the entities should be renamed to get standard names: synonyms (different designations for equipment tools, lots, and products), homonyms (equipment and chambers, process specifications and process specifications group). To bring some data together in a simulation input table, they were joined: lots and route steps to build a WIP file with current lot position; process and equipment information to build a route file etc. One of the typical problems of a real fab is a minor data inconsistency in data bases. Usually it is happening, because of human factor and has a very insignificant influence on the production process. Nevertheless a simulation model could not be run with incomplete data. That is why it was developed several approaches to reduce the influence of data inconsistencies in the real DB: it was taken a next operation for the lot, if the current operation was not found; the average process time for a process was used, if the process time for particular process was not found; dummy equipment was used if a particular equipment was not found. Today the aLTS model is used in Infineon Dresden. The application of the aLTS model is a dynamic capacity planning. It provides weekly forecast of several key performance indicators like cycle time, work-in-process, and delivery time for the next 3-4 months on fab, product group, work center levels. For example, in the Figure 7 a WIP-waves prediction report is represented. In this case WIP profile shows, that the most important problems are in segments 1...3 (left-top corner), but after 8 weeks the WIP-wave will move to segments 11-14. The simulation helps to identify most critical work centers and to trigger focused follow-up actions in production line.

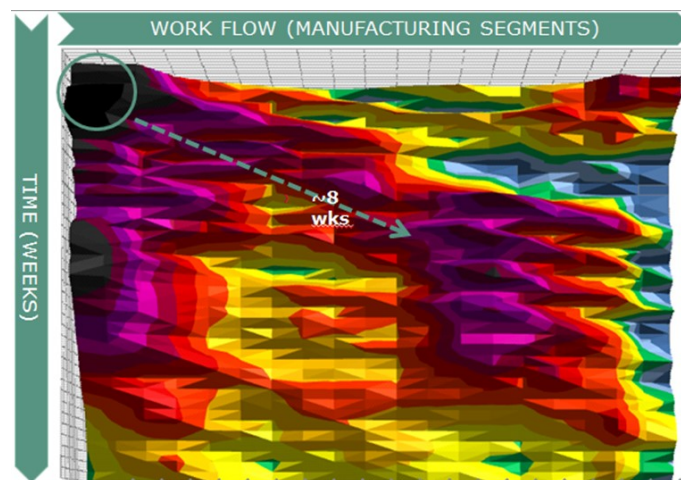


Figure 7: WIP-waves prediction report



### Conclusion

In this paper the results for the implementation of the simulation model building infrastructure for data analysis are represented. The core element is the automated-generated Long-Term Simulation (aLTS) model. The infrastructure allows building the aLTS model in an automated way, without manual steps in the extraction, transformation and load. During the implementation great efforts were made to meet sometimes contradicting requirements. Weekly meetings with industry experts have taken place for this purpose. The ideas and some results of the work were also approbated at two conferences (APCM 2017, 2018).

### About Authors

**Igor Stogniy** is a member of the scientific staff at the Chair of Technical Information Systems of the Technische Universität Dresden. He received his M.S. degree in Automotion and Control from Bauman Moscow State Technical University. His research interests include modeling, analysis and optimization of production systems.

**Wolfgang Scholl** works as a Senior Staff Expert for modelling and simulation for Infineon Technologies in Dresden (Germany). He studied physics at the Technical University of Chemnitz (Germany) and graduated in solid-state physics in 1984. From 1984 to 1995 he worked as a process engineer for ZMD in Dresden. In 1996 he joined Infineon Technologies (former SIMEC) and worked in the field of capacity planning. Since 2003 he is responsible for fab simulation. He supervises development and roll-out projects and is also a member of the Supply Chain Simulation community.

## 2.16 Web-Based Virtualization Toolbox for Integrated Visualization of Data

Author: Thomas Wagner; Technische Universität Dresden

Author: Germar Schneider; Infineon Technologies Dresden GmbH Co. KG

*To successfully manufacture the highly complex logic and power semiconductors in existing high mix semiconductor factories, an increasing degree of automation of logistic- and manufacturing steps is mandatory. But even in such highly automated environments, human interaction is still indispensable, e.g., to perform daily routines such as cleaning- and maintenance operations. However, the increasing amount of automation leads to a decrease in personnel working directly on the shop floor which is assigned to more and more tasks spread out to growing areas of the fab, thus limiting their access to KPIs and data required for the execution of their tasks. Therefore, methods for accessing the relevant data at the right time and place are required. For this purposes, a web-based virtualization demonstrator for fab data visualization is presented in this article which focuses on a combination of flexible use on a variety of end-user devices, modular dashboard design and the incorporation of modern web browser features. The main goal was to provide a toolbox containing a large array of possible visualization tools for different use-cases and areas of application.*

### Introduction

In the last years, the production spectrum of semiconductor manufacturing plants (fab) has changed in favor of a wide variety of customer specific, energy efficient products, especially in the logic market. Such a fab produces several hundred different, highly complex products. This complexity and variety leads to a large amount of logistic- and manufacturing steps that requires a high degree of automation to remain economically feasible. However, even in automated manufacturing environments, human intervention is still necessary. On the one hand, critical situations must be resolved, e.g., investigate sometimes transient and subtle process anomalies and deviations. On the other hand, daily routines such as cleaning- and maintenance operations must be performed. However, due to the fact that more and more production tasks are getting automated, the majority of staff requirement is shifted from the manufacturing- to the planning divisions. This leads to a decrease of the amount of personnel, as well as the available display work stations, in the field. The remaining personnel is assigned to more and more tasks spread out to growing areas of the fab while in the same time decreasing their efficient access to relevant KPIs and data, thus hampering their abilities. This can be seen as one form of the irony of automation. To assist the personnel during the execution of their tasks, methods for accessing the relevant data at the right time and place are required.

### Description

To solve the aforementioned challenges, a role- and ID- based visualization- and notification system was developed. The visualization of the data and KPI's is accomplished by mobile devices (ideally "hands-free") which also submit the current location, user role, user interaction as well as identification data of detected objects to a central knowledge based data management system. This system uses the combination of the aforementioned information, the "identification tuple", to classify the respective situation and calculates fitting parameters, KPI's or analysis- and diagnosis results. This data then gets transmitted back to the mobile clients where it is visualized. In the following, the development process, the technology as well as the current state of this demonstrator will be described.



### Overview

Initially, a collection of fundamental use-case information from the industrial partners was required. This included, e.g., the expected types of personnel roles, like maintenance engineers, operators or process engineers as well as the types of objects this personnel should be able to interact, i.e., to which data should be visualized. These include, besides production lots and manufacturing equipment, transport vehicles or even whole manufacturing areas. Additionally, it had to be defined what information should be available for which object, i.e., which use-cases are suitable for a first demonstrator. Finally, important discussions about the selection of suitable end-user devices and means for identifying the aforementioned objects, like QR-, RFID- or NFC-tags, were necessary. Especially privacy- and data security issues for W-LAN based localization and data transmissions had to be discussed. Based on these findings, several demonstrators have been designed, tested and finally integrated into the current version. The current version of the demonstrator is designed as a responsive web app, which enables a wide variety of end-user devices due to the cross-platform availability of modern web browsers. Its main visualization feature is a map-based main view. This allows for the representation of the core manufacturing assets, like production equipment or the automated material handling system, of a manufacturing site or even multiple sites (see Figure 1). The identification tuple is assembled using a conventional user management system including the user's role in addition with a system for object interaction.

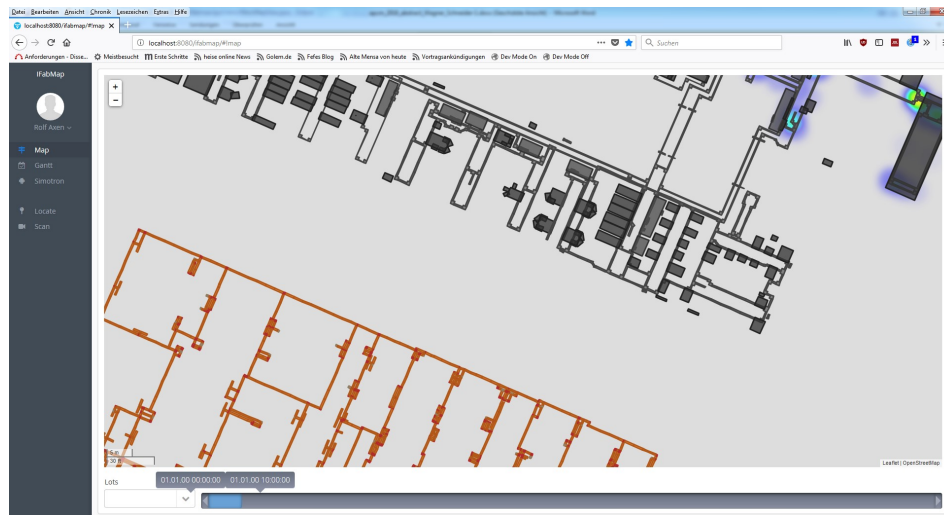


Figure 1: Main system view with multiple, disjoint manufacturing modules, partly featuring a heat map visualization

Currently, the user can choose between gesture based object selections, scanning labeled objects using the devices integrated cameras (QR-Codes, see Figure 2) and by utilizing the localization features integrated in modern Browsers (see Figure 3). To visualize the relevant data for different entities, a number of visualization methods have been designed. Depending on the identification tuple transmitted by the end-user device, data is either displayed directly on the main map view, specialized dashboards or both (see Figures 4 and 5). Currently, the visualization methods and dashboards are limited to data about productive equipment, transport system components and storage facilities. However, the amount of dashboards can easily be extended to integrate additional objects and use-cases. The visualization type and parameters is currently configured using JSON data files. This also applies to the available data collection and -preprocessing modules. Based on the currently defined use-cases, these modules provide versioned access to fab layout-, transport log- and FOUP vibration data, which is periodically assembled and updated.

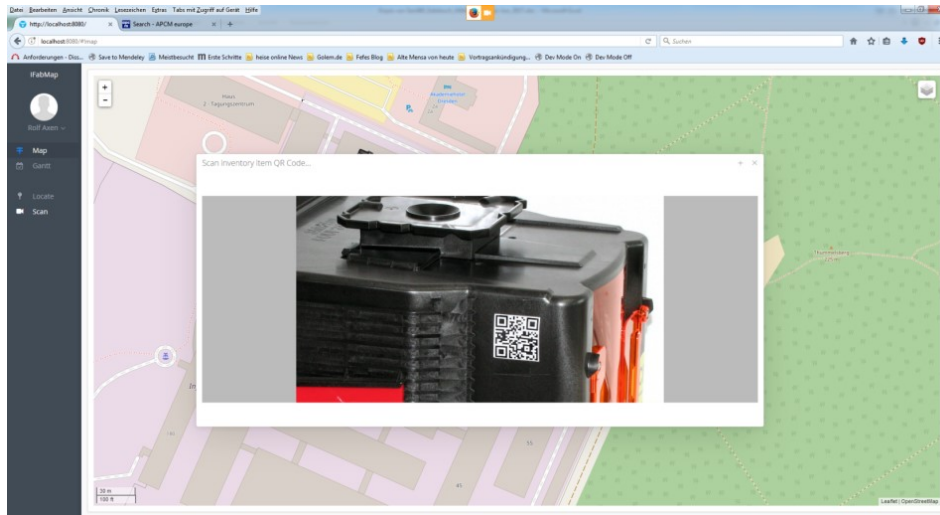


Figure 2: QR-based object selection using cameras integrated in end-user devices

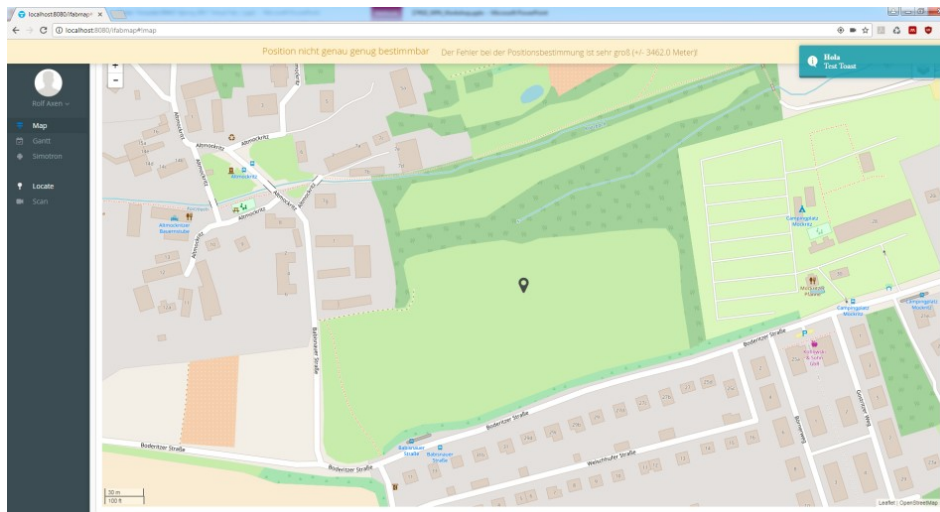


Figure 3: Integrated user localization using WLAN or GPS (if available, web browser dependent)

### *Implemented Technologies*

The demonstrator introduced above was built using a state-of-the-art web framework, the Vaadin Java UI framework [219], which greatly simplified the web app development for this demonstrator. The code is mostly written in Java, the HTML5-based rendering of the app as well as the automated communication between browser and the server is mostly handled by this framework. The framework already includes a wide variety of UI components. Missing functionality was incorporated by utilizing an add-on out of the add-on library provided by Vaadin [220] or implemented manually using the built-in, HTML and JavaScript extensibility. Examples for features that had to be manually developed during the course of the project are mainly in the field of using advanced features of modern web browsers. These components include:

- Geo location add-on: this script allows for an on-demand tracking of the users location to provide support for location-based task assignments or to provide directions (see Figure 1).
- A component to use the vibration functionality of mobile devices, if present. This can be used to signal the availability of data updates, new tasks, alarms or messages as well as to assist in user navigation.

- A component to access the devices web camera through the browser, in addition, a JavaScript library [120] to decode QR codes was embedded in this component for user- and object detection purposes (see Figure 2).
- A range slider component based on jQRangeSlider [62] to select ranges of numbers and time spans. This is, e.g., used to track lot movements or vibration data over time.

User authentication was accomplished by using the popular Java security framework “Apache Shiro<sup>TM</sup>” [7]. The responsive web front-end follows a two-layered approach. The first layer consists of the main menu and large view components, e.g., the widely used map view component or a Gantt-chart for task planning and visualization. The second layer is reserved for a dynamic amount of dashboards that can, at the moment, be configured using a JSON based, plain text configuration approach. This layer is managed by a server-based window management component that takes care of dashboard positioning, size restriction and window docking (see for example Figure 4). As mentioned before, the dashboards and other visualization methods are displayed depending on an information-tuple consisting of the current users ID, current location, user interaction as well as identification data of detected objects (QR code) or the object the user interacted with. Some of the information is optional within the configuration, e.g., an omitted user id will make certain visualization methods available to all users. In addition, a hierarchy (tree-based) structure of factory inventory is implemented in preparation for the new, knowledge-based back end system (see chapter 6). A simplified example is shown in Listing 1, where stockers and transport system elements are not referred by individual but by their class, resulting in the same visualization for all elements of this class. The dashboards themselves are an extensible set of Java classes that is enumerated dynamically at system start. A similar approach is currently implemented for the backend system. The data acquisition modules are again a set of Java classes. After enumeration at system start, the modules are scheduled one-time or at a fixed rate to look for new or updated data. The configuration of these modules, e.g., paths to data files or SQL connection information, takes place in a different JSON file, a simplified version is shown in Listing 2. This system is scheduled to be replaced by a more sophisticated approach as outlined in chapter 6. The data backend for storing new and updated raw data, already processed work units or buffer data is currently implemented using an ORM based approach through the widely used JPA framework EclipseLink [48]. This provides a certain degree of freedom with regards to the employed data store, currently a MariaDB [137] instance, as well as enables the data acquisition modules to define their own data entities (e.g. database tables) thanks to “dynamic weaving” [200].

## Results

As mentioned before, aiding the personnel of semiconductor manufacturing plants with customized, in the field visualization of data and KPIs is extremely valuable. By using a responsive web based system as the one demonstrated in this contribution, a significant degree of freedom can be achieved since web browsers are available on a wide variety of devices ranging from Smartphones to desktop PCs or even large monitors on the shop floor, providing a consistent look and feel. The features of modern web browsers can also be used to implement advanced localization and data acquisition methods to enable the users to quickly obtain the required information about an object or a task at hand without the need of a cumbersome search in traditional input forms. Additionally, a lot of useful information can be displayed with the use of compact dashboards and map-based data visualization methods as shown in the proposed software demonstrator. Although these methods are suitable for most devices, some of them, e.g., data glasses may feature even smaller screen sizes that restrict the representable information to a very small amount of text and/or colors. The abovementioned identification tuple must therefore be extended by information about the end-user device, introducing another selection layer for the backend system and inflating the JSON based configuration infrastructure. Therefore, current research work includes moving away

from this system towards a self-description of the visualization-, analysis- and data processing modules drawing on a knowledge-based description of the manufacturing system. This enables the automated assembly and validation of visualization- and data analysis module chains that may even involve sophisticated simulation components to enable predictive KPIs.



Figure 4: Visualization of AMHS- and storage data using dashboards and a map based status display

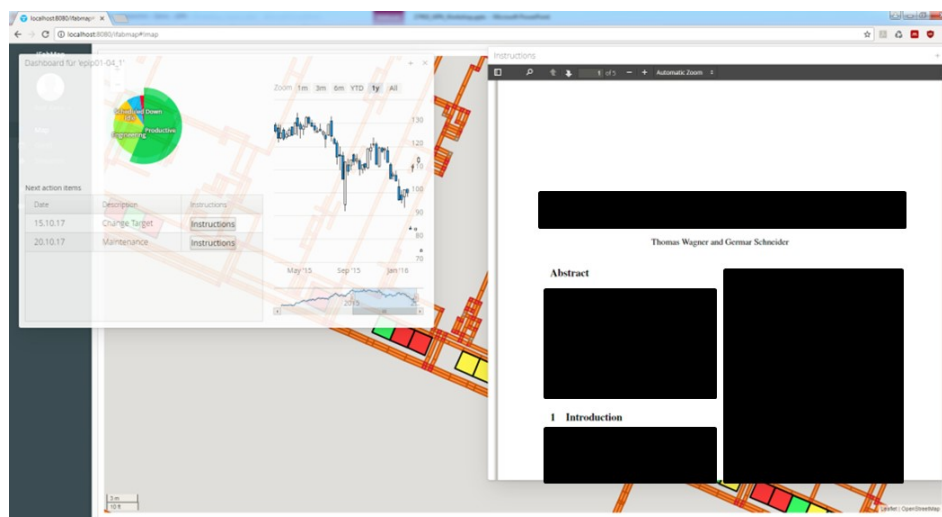


Figure 5: Visualization of equipment data using a dashboard and a vision of a web-based manual- and reporting display (based on pdf files)

**About Authors**

**Thomas Wagner** is a member of the scientific staff at the Chair of Technical Information Systems of the Technische Universität Dresden. He received his M.S. degree in computer science from Technische Universität Dresden. His research interests include modeling, analysis and optimization of logistic systems as well as the automated diagnosis of production equipment.

**Germar Schneider** holds a Diploma in chemistry and a PhD in analytical chemistry of high purity materials. He joined the Siemens AG in 1995 in the joint venture of SIEMENS/IBM in Essonnes and worked for three years as a process engineer in the Wet department. In 1998 he changed from Essonnes to Dresden and was responsible as a section manager for the 200 mm Wet department. From 2004 to 2008 he installed a team for factory automation integration, which was a basis for several upcoming factory and automation projects. Between 2008 and 2012 he worked as a manager within the Infineon wafer test department founded in 2008 and was responsible for production equipment engineering. With more than 20 years of experience combining know how of process engineering, production, maintenance and automation he worked as a work package leader within the EU-projects EPT300, EPPL, SemI40, Productive40, and iDev40.

## 2.17 Tasks and Results of Work Package "Swarm Intelligence"

Author: Karli Hantzschmann; Fabmatics GmbH Dresden

Author: Germar Schneider; Infineon Technologies Dresden GmbH Co. KG

*Self-Navigating Systems (SNS, sometimes also called AGVs) promise to be an interesting and new approach to material handling automation in semiconductor fabs to stretch the limits of existing traditional transport systems. This article briefly describes the tasks and results of the work package "Swarm Intelligence", which was formed with SEMI4.0 to tackle some of the most important challenges of this new automation technology the consortium faced in SEMI 4.0 project.*

### Introduction

One potential solution to overcome the flexibility limits of standard transportation systems like Overhead Hoist Transport Systems (OHT) or Conveyor Systems are Self-Navigating Systems, which are able to move and handle the highly sensitive payloads (not only wafers, but also reticles, probe cards, spares and wears, etc.) autonomously. The absence of any mechanical infrastructure like rails, markers, stripes etc. gives these systems an unbeatable flexibility in terms of layout and performance scalability because they can be adjusted at any time to match the changing production environment without any re-construction efforts, which makes them the perfect companion to master the challenges of industry 4.0 in semiconductor fabs. Although first SNS have been developed to the stage of industrial application within previous funded projects like CoolFab or EPT300, research and development was focused on the capabilities of the individual systems, whereas the simultaneous application of multiple systems at one production site (i.e. within a "swarm") creates new challenges which need to be understood and tackled within this project.

### Tasks and Work Package Structure

The following figure shows the three focus areas the work package activities were focused on:

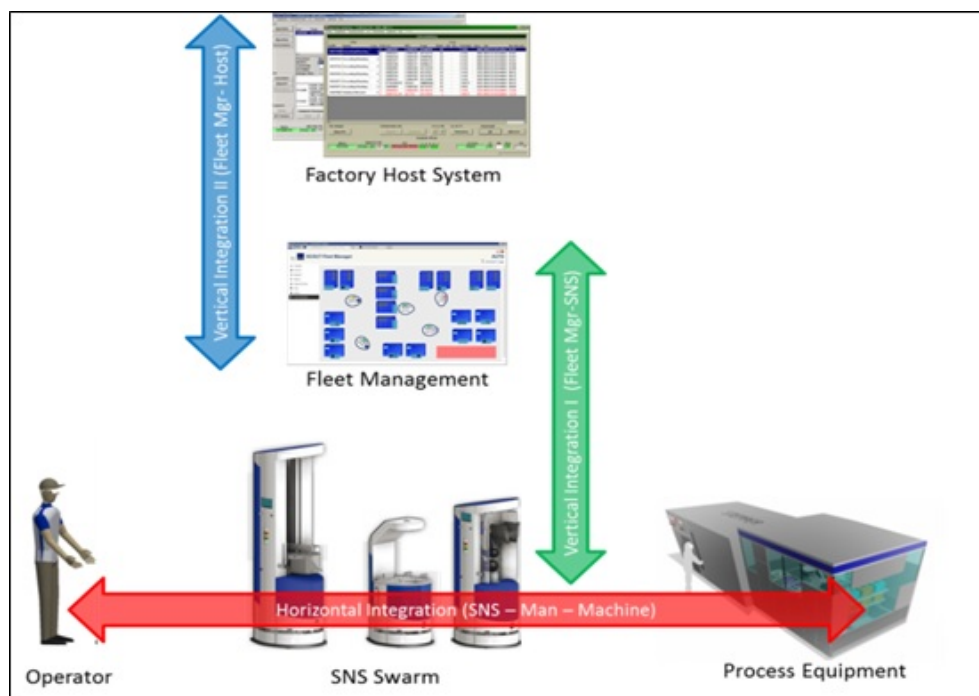


Figure 1: Focus Areas of work package „Swarm Intelligence“



The main goal of work package „Swarm Intelligence“ was to understand the complexity arising from using multiple SNS (of same or different type) in parallel within the same area. For this reason a new, highly flexible SNS platform and solutions for control, tracking, communication and collaboration of a swarm of SNS within a 24x7 production environment under highest cost and quality pressure had to be developed – aggregated under the keyword Swarm Intelligence. The results of the work package were supposed to be demonstrated within a use case called „Demonstrator of a swarm of SNS“. In order to successfully run the project and to implement the aforementioned use case the work package was split into the following main tasks:

1. design and construction of an advanced mobile robot platform
2. development of a robust vehicle-to-vehicle communication to prevent collisions and to allow a proactive obstacle avoidance
3. concept, design, implementation and test of an intelligent fleet management system including development of an appropriate host interface allowing customers to interact with the fleet of multiple mobile robots

### Main Project Results

This section gives an overview about the most important results of work package “Swarm Intelligence” reached by Fabmatics in close cooperation with its main partners MetraLabs, Infineon and Bosch.

#### *New mobile robot platform x3*

After a detailed requirements collection and analysis phase with all potential customers Fabmatics and MetraLabs developed the concept of a new and highly flexible mobile robot platform (SCOUT X3). Due to its rectangular shape the new SNS platform provides both significantly improved space efficiency and mechanical stability compared to its predecessor’s generation. The modular and scalable platform concept allows to flexibly adapt the system to various customer requirements, such as different payload sizes and weights, various buffer capacities or different automatic handling mechanisms. A new generation of internal drive and control components such as safety motor controller, battery management unit, encoders or control PC enables a higher system performance with better navigation and docking behavior and greater precision of the 3D obstacle detection.

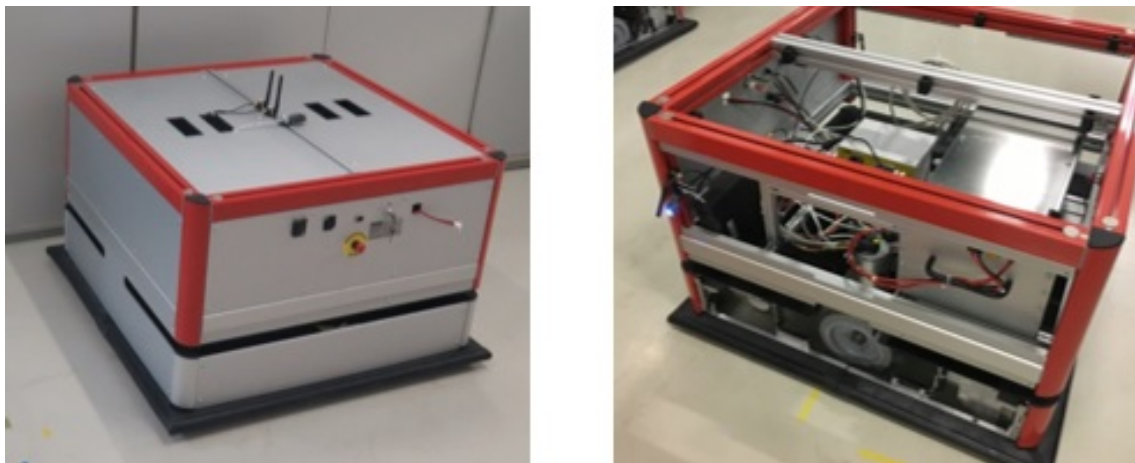


Figure 2: Prototypes of the new SCOUT X3 SNS Platform

For automatic material pick and place the platform can be equipped with different handling components, such as conveyors or standard 6-axis industrial robots with various grippers.



### *Robot-to-robot communication*

As long as only one SNS was acting in one area at one time there was no obvious need for a reliable robot-to-robot communication. When using swarms of multiple autonomous systems simultaneously in shared areas the risk of inefficiency by time eating obstacle prevention or even collisions between different vehicles becomes relevant. For this reason a direct robot-to-robot communication mechanism based on ZigBee was developed, implemented and successfully long-term tested within the production facility of a Bosch. The following picture illustrates the principle by showing the map of a particular SCOUT (here marked green). Another SCOUT (yellow) and a HEROfab SNS (also yellow), which is shaded by the other SCOUT, are fully visible due to ZigBee communication and therefore can be included in the proactive route planning to avoid deadlocks and collisions.

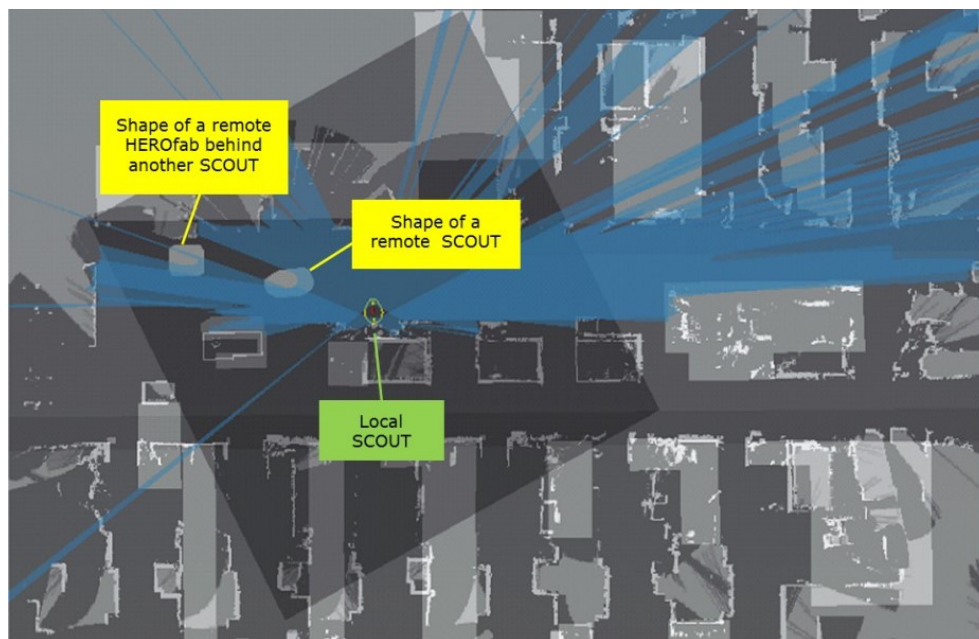


Figure 3: SCOUT real time map showing two other vehicles visible due to ZigBee communication

### *Hero scout fleet manager (HSFM)*

When multiple mobile robots share common areas an efficient and flexible control systems becomes mandatory. Together with all project partners at first a detailed requirements collection and engineering was performed, which formed the cornerstone for the concept and the design of a new fleet management system called HERO SCOUT Fleet Manager (HSFM). The new fleet manager should be able to efficiently control homogeneous or heterogeneous fleets of mobile autonomous systems, not only Self Navigating Systems like SCOUT or HEROfab, but also rail guided systems (RGVs) such as Fabmatics' HEROfario. Two main modules represent the functional heart of HSFM and are responsible for

- vehicle dispatching, i.e. the mapping of incoming transfer requests onto the best suitable mobile system, and
- vehicle routing, i.e. controlling all vehicles in real time to move them on the best available routes from their current locations to the requested destinations (and preventing deadlocks with other vehicles)

The HSFM internal vehicle dispatching function uses an efficient game tree search based algorithm. It is implemented within a pluggable module which allows easy changes and adjustments based on specific customer requirements. Beside the two core functions HSFM additionally implements

functions

- host interface
- real time graphical user interface (GUI)
- configuration and administration
- alarm management
- material and transfer job management
- automatic vehicle charging
- exception handling
- advanced logistical functions like one-way streets, no-go-areas, speed zones, etc.

The figure below shows the general architecture of HSFM.

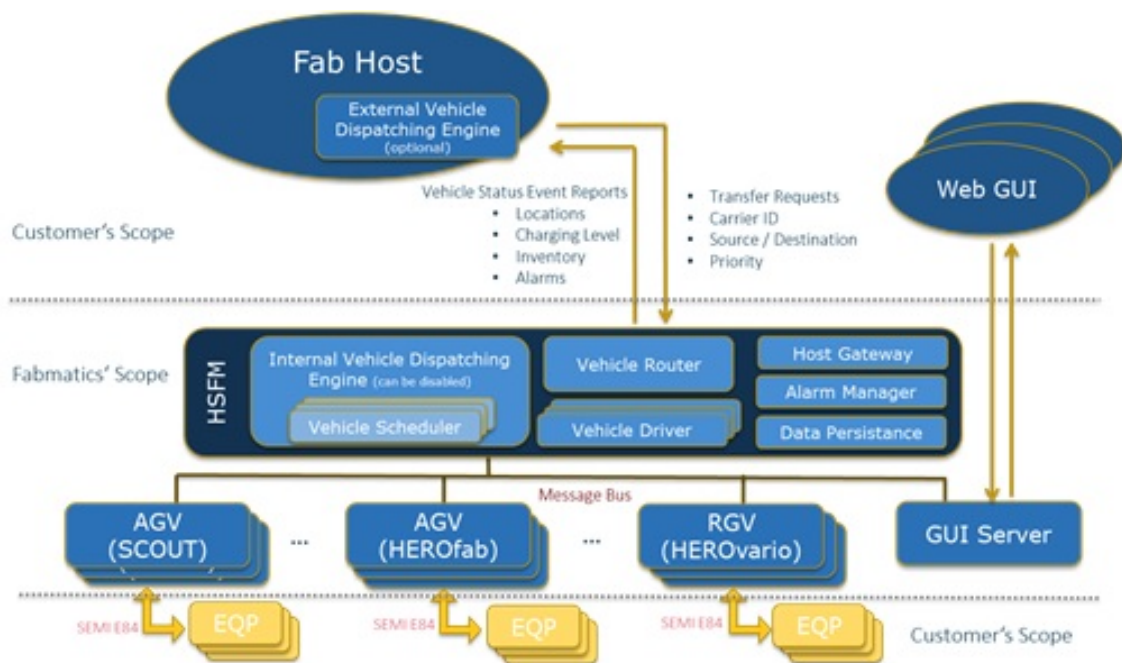


Figure 4: General HSFM Architecture

Apart from the fleet manager's internal functionality it is important for each customer to have a powerful, standard-compliant and flexible interface allowing the fab host system to interact with the SNS fleet. Within the work package a SEMI standard based HSMS/SECS-II interface was designed and implemented by Fabmatics. For customers like Infineon it was very important to have the possibility to develop and implement own customer specific vehicle dispatching mechanisms. For this reason Fabmatics developed together with its partners Infineon a concept for an optional external vehicle dispatching engine (VDE) which can be used instead of the HSFM's internal vehicle dispatching engine (which must be disabled in such a case). Using the external VDE requires a different host interface concept. The following figure illustrates the concept of optional external and internal VDEs.

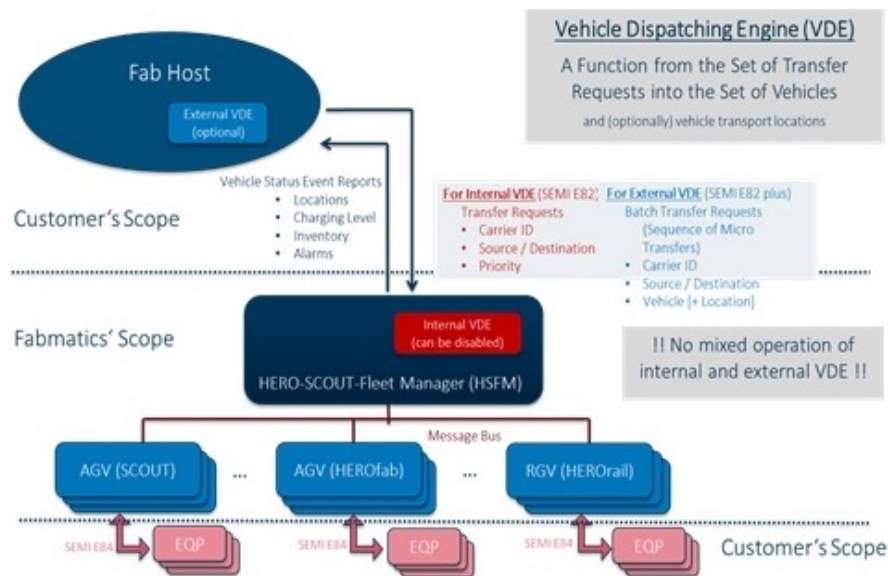


Figure 5: Concept of internal vs. external Vehicle Dispatching Engine

### Use Case Demonstration

The results reached by Fabmatics and its partners as described in the chapter above formed the base for the planned use case demonstration. Based on the new X3 SNS platform a new mobile robot SCOUTactive neo with automatic material handling was successfully developed. The system is equipped with 6-axis Schunk Powerball industry robots with customer specific gripper. As a test area three types of dummy process equipment with multiple loadports had been installed. 17 cassettes were automatically moved between the different source and destination ports by a fleet of 4 SCOUTactive neo SNS, controlled by the newly developed HSFM fleet manager, which was connected to a dummy host system. The fleet manager successfully managed the parallel operation of the fleet of 4 SNS in the narrow test area by using its internal VDE and its sophisticated vehicle routing algorithm. The web based real time GUI displaying the used routing map is shown in Figure 6 below. The operation of the intelligent fleet could be successfully demonstrated at Fabmatics site in Dresden in August 2018.

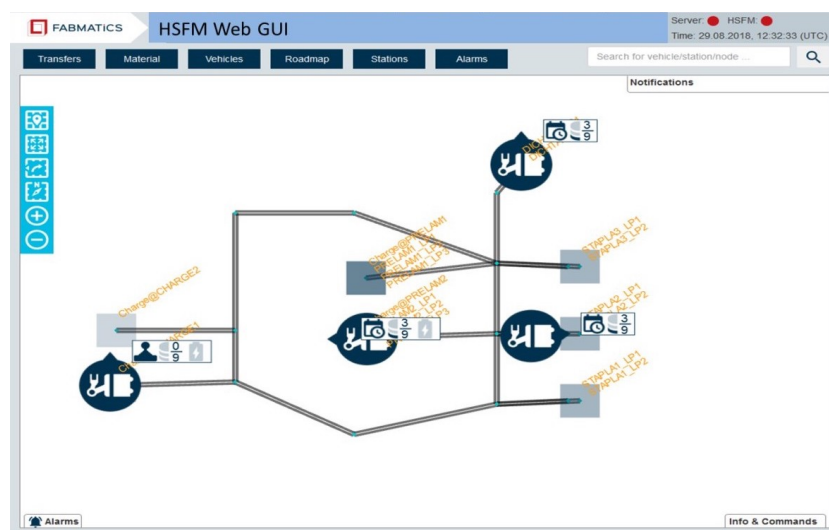
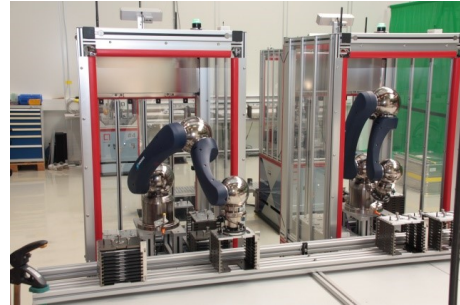


Figure 6: Web based real time HSFM GUI

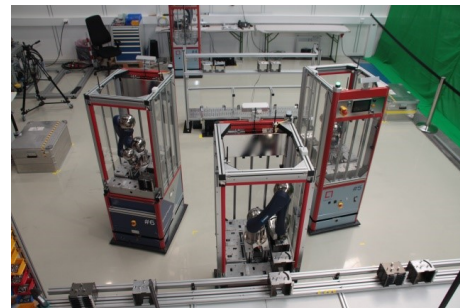
Photos below show SCOUTactive neo SNS and give an impression of the use case demonstration event.



(Figure 7:) SCOUTactive neo handling customer specific cassettes



(Figure 8:) SCOUTs during cassette hand-over to equipment



(Figure 9:) SCOUTs during test run - 4th SCOUT is charging in background.

By the use of autonomous robotic transportation systems lead and transport times can be reduced and, thus, the overall throughput and utilization of the production line can be increased. This, in turn, leads to an increased automation level and improved productivity of the semiconductor fabs. In addition, new optimization potentials will arise in those production areas that have already reached the limits of profitability possible with the limited automation technologies available until now. The use of mobile robots supports the operators in case of physical strain or even relieves the completely of certain tasks, thus freeing up time for more value added activities.

### About Authors

**Karli Hantzschmann** Hantzschmann studied mathematics at the Technical University of Dresden and holds a PhD degree in computer science. He joined Siemens Microelectronics Center in 1998 and had been working in fab automation for 11 years. In 2009 he changed to Roth Rau - Ortner GmbH and was first responsible for worldwide solar turnkey business service. In 2011 he became business unit manager for Ortner's fab automation business. Since 2017 he has been acting as the department manager for project management of Fabmatics GmbH. In his role he is responsible for the management of all projects of his company.

**Germar Schneider** holds a Diploma in chemistry and a PhD in analytical chemistry of high purity materials. He joined the Siemens AG in 1995 in the joint venture of SIEMENS/IBM in Essonnes and worked for three years as a process engineer in the Wet department. In 1998 he changed from Essonnes to Dresden and was responsible as a section manager for the 200 mm Wet department. From 2004 to 2008 he installed a team for factory automation integration, which was a basis for several upcoming factory and automation projects. Between 2008 and 2012 he worked as a manager within the Infineon wafer test department founded in 2008 and was responsible for production equipment engineering. With more than 20 years of experience combining know-how of process engineering, production, maintenance and automation he worked as a work package leader within the EU-projects EPT300, EPPL, SemI40, Productive40, and iDev40.



## 2.18 Closing the Gap Between Exploding Amounts of Data

Authors: Gerhard Luhn; SYSTEMA

Authors: Johannes Postel; SYSTEMA

Authors: Lutz Muche; SYSTEMA

Author: Philipp Kopp; Elmos

Author: Germar Schneider; Infineon

*The need for real-time Big Data Analytics impacts semiconductor industry. Companies steer via digital twin and smart manufacturing (Industry 4.0) into an intelligent enterprise. However, as emerging IoT devices will create zettabytes of data, a new quality of real-time data-transformation-into-information is required. “73% of semiconductor executives believe that many companies have not yet invested in the capabilities to verify the truth within the data used in their most critical systems” [188]. This paper describes a new, logical-physically inspired approach to meet those challenges. In order to enable real-time decisions within ever growing amounts of data, companies need to become more flexible, liquid and scalable. Within the traditional mindset of data warehousing, data modeling does not have a logical or physical foundation, but remains as an “art” [52]. Data structures are created based on experiences. Consequently, the complexity of data mining seems to be undecidable, and thereby creating a source for disambiguation and pain points in data processing [52].*

### Introduction

Scope of this paper is to thoroughly sketch a methodology providing most powerful algorithmic efficiency, solving the algorithmic complexity issue while transforming each data into physically and semantically clear informational spaces. Those spaces are enabling lowest algorithmic complexity, as based on inherent informational structurization [132]. This inherent structurization lays down in detail, how specific emergent behavior of complex production systems (such as: dynamically moving bottlenecks, propagation of material waves, etc.) are always – in a holistic / informational manner – compositions of atomic events. However, such emergent pattern may create own lawfulness – similar to fluid phenomes as described by Navier-Stokes equations.

In one word: if we ground the data on an informational holism, and make the computing system thoroughly scalable and decidable. This enables

- a) decisions of your experts as based on real-time and clean information, and
- b) to create an overall information system and architecture as based on most efficient data structurization, guaranteeing decidable and minimal algorithmic complexity.

The potential of highly efficient information processing in semiconductor manufacturing will be explained. A (quantum-)physically inspired holistic information concept will be introduced, bringing new types of solutions into the foreground.

This new approach has been initiated by Systema, and has been further developed within the Semi40 project together with partners Elmos and Infineon Dresden establishing a demonstrator resulting in a software platform which can be used for fast and reliable data analytics to many different key performance indicators in a semiconductor fab with a highly complex product portfolio.

### **The Growing Gap Between Data and Information**

There is a growing gap in our understanding of the world, and especially in engineering science. There is a physically described world, including manufacturing systems, which is the digital twin in semiconductor manufacturing. With increasing automation level of the fabs the digital twin becomes more and more relevant, and therefore the managing of the data is very important and will be the key to enhance the productivity.

Today's semiconductor fabs have already big computing systems and server systems with high storage capacities, including factory control and automation systems. However, while information processing in nature follows the principle of minimum energy (least action [215]), computing does not. Nowadays huge effort are spent to optimize energy efficiency in computing [49], [190]. Additionally, the truth- or meaningfulness of data becomes a growing problem. Intel collects over 5B sensor data-points per day per factory and it's growing: 100's of points per second per sensor collected, with 1000's of sensors throughout a factory [33].

Semiconductor manufacturing produces huge amounts of data, and this is even enforcing within the IoT-decade. Parallel to this the upcoming more-than-Moore technologies are bringing the production sites on new levels of uncomfortable complexity. Decisions need to be taken in real-time, and new functionalities are required, like continuous production forecasting (ex.: continuously forecasted detection of production bottlenecks).

In recent decades, first applications and software solutions have been developed, in order to enable highly automated semiconductor manufacturing. Unfortunately, many of those systems lack data / information consistency, speed, user-friendliness, flexibility and comprehensibility. Furthermore, the underlying data heterogeneity has steadily increased and has led to a high level of system complexity that no longer meets the current requirements of a digital factory.

Additionally, classical methods of software engineering do not support a conceptualization of an isomorphism between physical reality (i.e. of manufacturing systems) and computing systems. While huge efforts are spent to increase power of hardware systems, and to create architecturally and functionally well designed computing systems, the level of algorithmic complexity remains unclear, and may cause "BI and ETL Process Management Pain Points" [18] (BI: Business Intelligence, ETL: Extract, Transform, Load).

To summarize, the informational gap opens up gravely. In order to solve these shortcomings, a new holistic information model (HIM) has been evaluated and further developed within the Semi40 project. The functional scope, performance and scalability of such new systems needs to be raised by orders of magnitude.

### **Toward a (quantum-) Physically Inspired Concept of Information – the Holistic Information Model HIM**

The starting point is a concept known in physics and systems theory as phase space. A phase space describes all possible states of dynamic systems. In the technical system to be considered here, this includes: system components, their states, and all events that lead to changes in the states. It is important for the dynamic systems (production systems) to be considered here that information on system states are generated using KPIs (cycle time, flow factor, operating point, etc.). In [132] (Luhn 2017) it is shown that all this information can be generated by aggregating simple informational components. This guarantees maximum algorithmic efficiency and fastest response times even for very complex user queries.

In computer science it is shown that there is no general method to determine which algorithm holds the lowest complexity. Therefore, solutions are often implemented whose complexity cannot be managed - especially with large amounts of data. However, the proposed concept guarantees the lowest complexity for all possible queries and thus the best possible effectiveness by linear summation.



How can this be achieved? What does it mean to thoroughly control the algorithmic complexity of a computing system, and knowing at the same time that users might show up and querying statements / algorithms of unknown complexity?

The key point is that the informational framework by itself (similar to knowledge domain in physics, for example mechanics) implements a physically clean and rich phase space, which will by definition of the informational space deliver responses to any meaningful query within best algorithmic efficiency.

The physicist Richard Feynman once mentioned that any particle selects among the millions of possible paths the true path, and this true path is the one, which consumes the lowest amount of energy (the principle of least action [215]). In another context he claimed that any physical or natural object like a coil works similar to a parallel computer, because it exhibits in real-time continuously answers to complex questions (for example: what is the magnetic flux of a phase coin within coordinates in the space-time continuum).

That is, any physical law reduces the entire phase space to an absolute minimum of possible descriptions. To turn it around: if a phase space is setup in a manner, that any input parameter (that is typically the question or the query to be executed) will be treated not by heuristically developed algorithms, but by law-like structures, then the algorithmic complexity of the system will be minimal, and each query will be resolved within an absolute minimum amount of steps.

What does that mean in terms of the physical framework? Study [132] outlines that if the systems are described in terms of a decompositional system model, then an informational model based on a decompositional system model creates mathematically linear spaces. That is, the superposition principle holds.

The superposition principle states that for all linear systems the net response at a given place and time caused by two or more stimuli is the sum of the responses which would have been caused by each stimulus individually. So that if input A produces response X and input B produces response Y then input (A + B) produces response (X + Y). The homogeneity and additivity properties together are called the superposition principle. A linear function is one that satisfies the properties of superposition. It is defined as

$$\begin{aligned} F(x_1 + x_2) &= F(x_1) + F(x_2) && \text{Additivity} \\ F(\alpha x) &= \alpha F(x) && \text{Homogeneity} \\ &\text{for scalar } \alpha. \end{aligned}$$

This principle has many applications in physics and engineering because many physical / engineering systems can be modelled as linear systems. This principle has been applied for the concept of holistic information. A simple example may be a falling ball. Its falling distance, time, speed etc. may be measured, and may create data. Now, the solution provides to any desired, even complex interrogation the corresponding answer within lowest algorithmic complexity. For example: a) which ball during the last years was the fastest? Which medium speed, and speed variability did he show? What was the fastest grouping of balls (with regard to the different groups of material the balls are made of)? The variability of speed within each grouping, and between the groupings? Some balls may have split of, or merged together – show the corresponding figures for all subparts of those balls. Other balls may have been thrown into liquids (instead of air). What about costs? The system may also show nonlinear behavior. The concept of re-entering material / lot flows in semiconductor manufacturing is an example. Never the less, the information system covering such nonlinear behavior has to behave in linear complexity – which is fulfilled.

The big advantage for semiconductor fabs e.g. IFD and Elmos is, that a fast computing and fab analysis using the holistic modelling will be a very nice tool not only for the engineers working in manufacturing but also for a lot of other customers working on fab analytics or machine learning, see Figure 1 below.

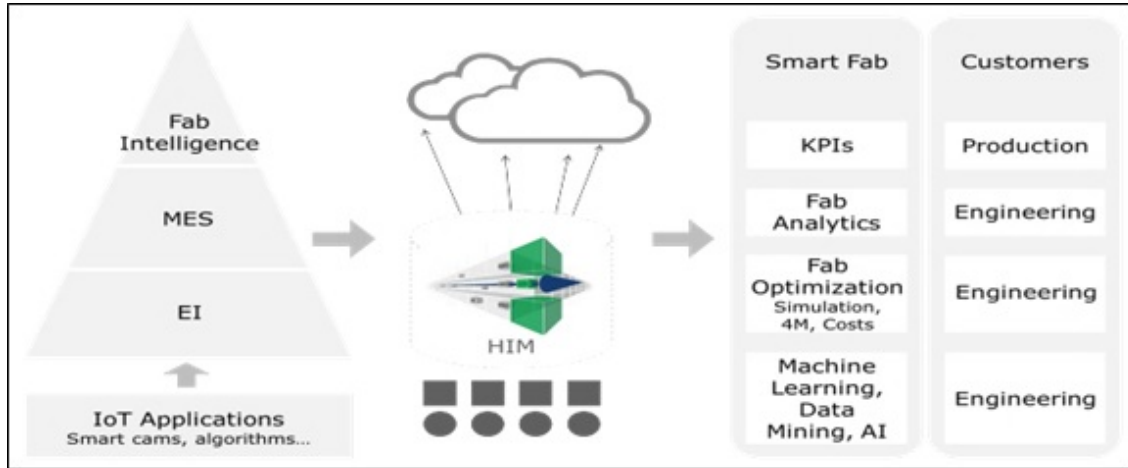


Figure 1: The HIM offers many customers in a fab KPIs and data analysis in a smart fab; see [183]

Within a smart fab a lot of new possibilities to improve the manufacturing process are shown from simple fab analysis to machine learning up to first approaches for artificial intelligence.

As the system is based on quantum physical principles, it is worthwhile to incorporate ideas about quantum computing as well. Quantum computing enables to process many potential threads in parallel. However, a quantum computer returns a single correct result, while a classical parallel computer directly returns all valid results. This is especially required to create informative overview sheets, scaling from single atomic events to elaborated forecasts of complex systems (all kinds of dashboards and data analysis platforms).

Quantum computing has interesting concepts; however, it may take 10 years from now to have a system covering a few thousand qubits running [71]. Let us analyze a simple, but highly representative scenario and doing a first assessment of quantum computing. Typical data warehouse schemes cover histories of different kind of objects (lots, equipment, transporting units etc.). A simple interrogation to calculate for example a cycle time or a flow factor may typically initiate two requests toward the database. The first request may ask for the starting timestamp, and a second request may retrieve the ending timestamp. Now, within a holistic model, only one request is required. We may also imagine to retrieve within a more complex query hundreds or thousands of timestamps. The corresponding algorithmic complexity for the classical history table grows with quadratic order (which is notes  $O(n^2)$  in computing science). Contrary to this, the algorithmic complexity for the holistic model – because it is linear – grows with linear complexity (which is noted  $O(n)$ ).

From here a first comparison to quantum computing can be drawn. Grover's algorithm is known to retrieve data out of data sets. For single interrogation its complexity equals the square root of the number of lines ( $n$ ) in the table:  $O(\sqrt{n})$ . What does that mean with regard to our scenario? If we use classical database schemes (histories), then the resulting complexity is  $O(\sqrt{n}) * O(\sqrt{n}) = O(n)$ . To summarize, the holistic model, even while using classical, von-Neumann computing architecture holds the same degree of complexity as quantum computing.

That is, this method gives you orientation about the entire quality of your overall data system, and enables you to efficiently manage and continuously improve this quality.

Never the less, future research will evaluate further details in how far quantum computing systematics might already be useful within this context. The next picture summarizes the outcome of this discussion.

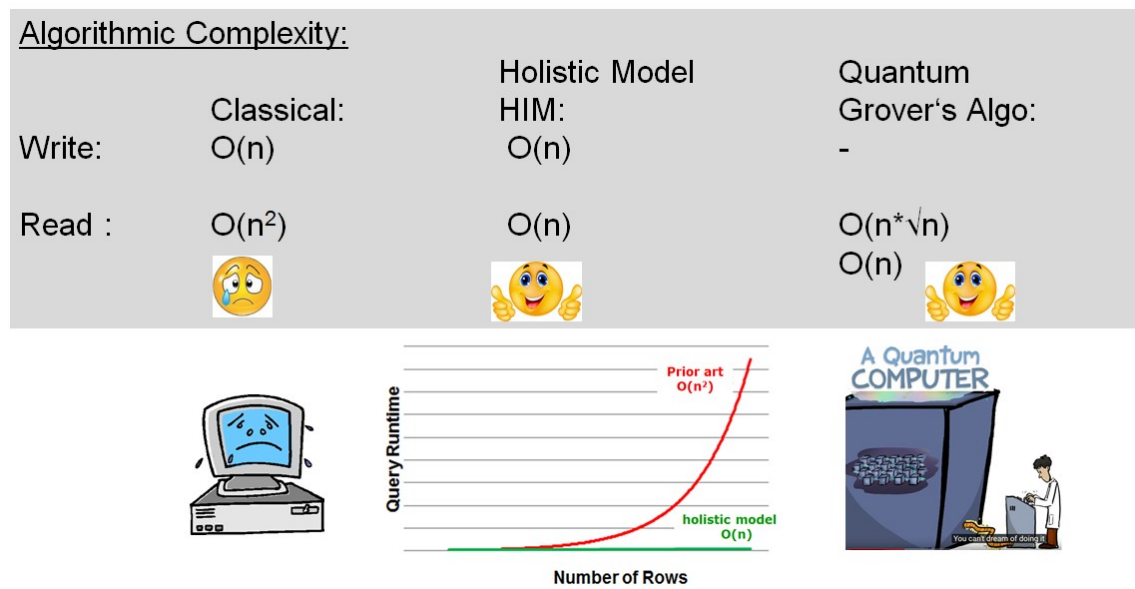


Figure 1a: Comparison of algorithmic complexity within 3 different contexts: a) von Neumann computing, classical data warehouse scheme (left); b) von Neumann computing, classical (red) versus holistic (green) information model; c) quantum computing, classical data scheme

Figure 1a shows an aggregation run, in one example over 10.000 data sets. The classical, batch-oriented solution (red color) requires 900 sec. runtime. The Real-Time aggregated solution requires 8 sec. runtime. This indicates a runtime reduction of factor hundred.

The main topic and key innovation of this concept of information is that it enables to overcome the dilemma of being drowned in oceans of data by design of a corresponding architecture. While the algorithmic complexity grows for state-of-the-art systems within quadratic order, the holistic model HIM grows with minimal algorithmic complexity (linear complexity in this example).

The holistic model provides also a solution or guidance to enable best truthfulness of data. While classical concepts extract, transform and load data from different sources for a specific period into the data warehouse (those mechanisms are called ETL processes), the holistic model turns this approach upside down. Here, any event gets mapped in real-time against the meta-model of the production process (which owns the truthfulness of how to transform data into information), and from here on the informational spaces are immediately populated. Additionally, errors are immediately recognized, and actions can immediately be taken in order to correct such errors. This concept is called SPOT “single point of truth”.

In other words, the pain points of ETL process management [18] will be systematically eliminated, and transformed into structural advantages of the new system and method.

### New Functional Application: Continuous Production Forecasting in Real-Time

Operative systems (such as MES, equipment integration, SPC) create data in real-time, which is now ready to be used for more advanced analytics, such as simulation and forecasting. However, data models for simulation are made for specific scopes, and are not easy to be integrated and maintained. Additionally, the actuality of a simulation run expires soon. Any new, non-simulated event may decrease the truthfulness of the simulation results.

Consequently a new requirement appears, covering continuous production forecasting in real-time. This can be imagined similar to weather forecasting. While starting this application, a weather forecast will be brought up, and will be continuously updated. One scope of such a requirement is to early detect upcoming production bottlenecks. The next figure illustrates an example of such a forecast functionality. Typically, products are produced along production routes, which are made of sequences of operations (called “Oper1, Oper2 ...”).

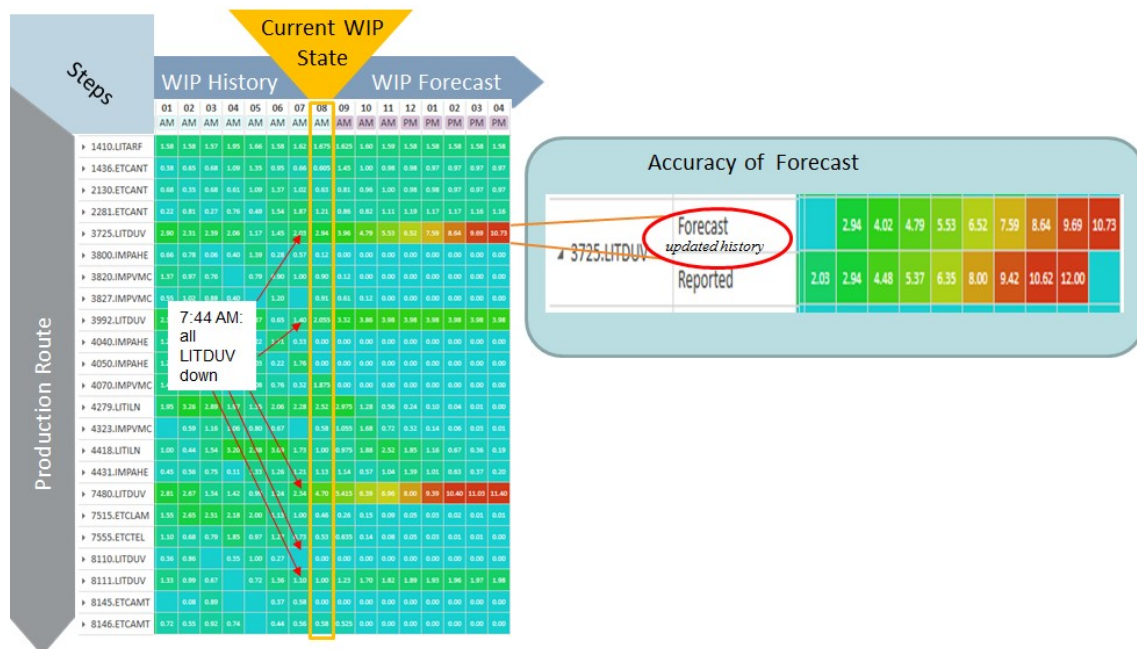


Figure 2: Holistic Production Forecasting HPF: example from a simulated Fab

Figure 2 shows a forecast of 8 hours. In this example, a specific incident happened: 7:44 AM: all LITDUV stepper went down (no point-in-time for recovery was specified). The forecast was immediately updated. This is an example, where the fab was simulated by a simulation tool. The right part of this figure shows two rows for the stepper 3725.LITDUV. The upper row shows the forecast, and the lower row shows how the material flows in real. In this case, the difference between forecast and real was around 20%.

However, production units may show high variability. Such a high variability is caused by different use scenarios of those equipment (different products running over the same equipment), and by natural variability. The next figure shows the difference of the production forecast in comparison to the real values (data provided by Elmos). Here, the difference is sometimes greater than 40%.

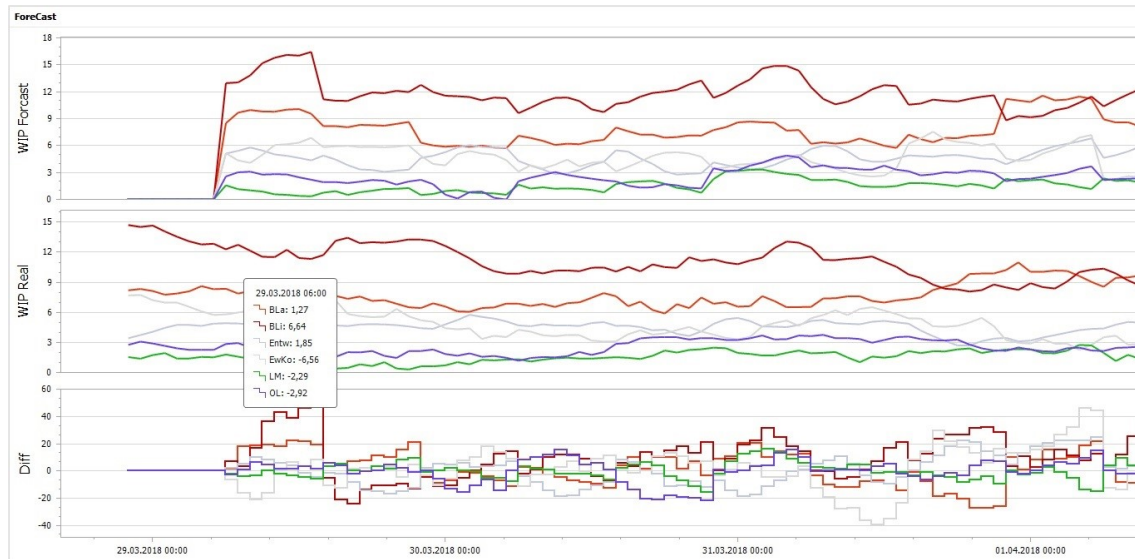


Figure 3: Comparison of forecast data with real data (Elmos production data): differences go up to 40%

Based on those results, a deeper analysis has been performed, in order to show on which critical factors the quality of production forecasting depends on. Following systematics has been applied. Three different Fab simulation model have been built, each adding more complexity to the production process. The model covers some 30 succeeding operations and representing a typical semiconductor production sequence.

Model 1 covers a linear process flow (equipment might be used several times in this flow). However, each equipment process behaves strictly deterministic, covering zero variability in process time. In addition, the start of new lots takes place in equidistant time slices (each 3 hours).

Model 2 adds a certain variability to the starting process of new lots (exponential distribution of starting times).

Model 3 holds the same characteristics as Model 2, but adding additional, statistically distributed tool downs into the production process. (in nächster Version / Abschlussbericht)

The next figure shows a succeeding sequence of operation steps (5 steps). The first row shows the starting operation. Then, each two row shows the forecasted data, and the real data.



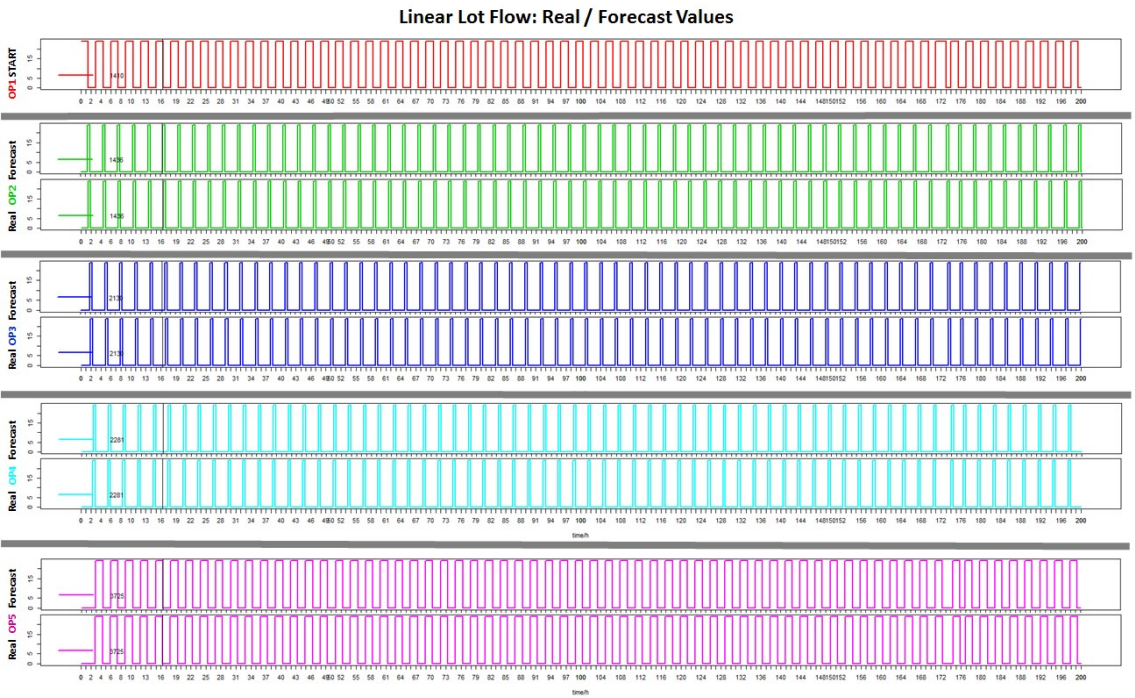


Figure 4: Comparison real values / forecasted values; most simple linear lot flow (no variability)

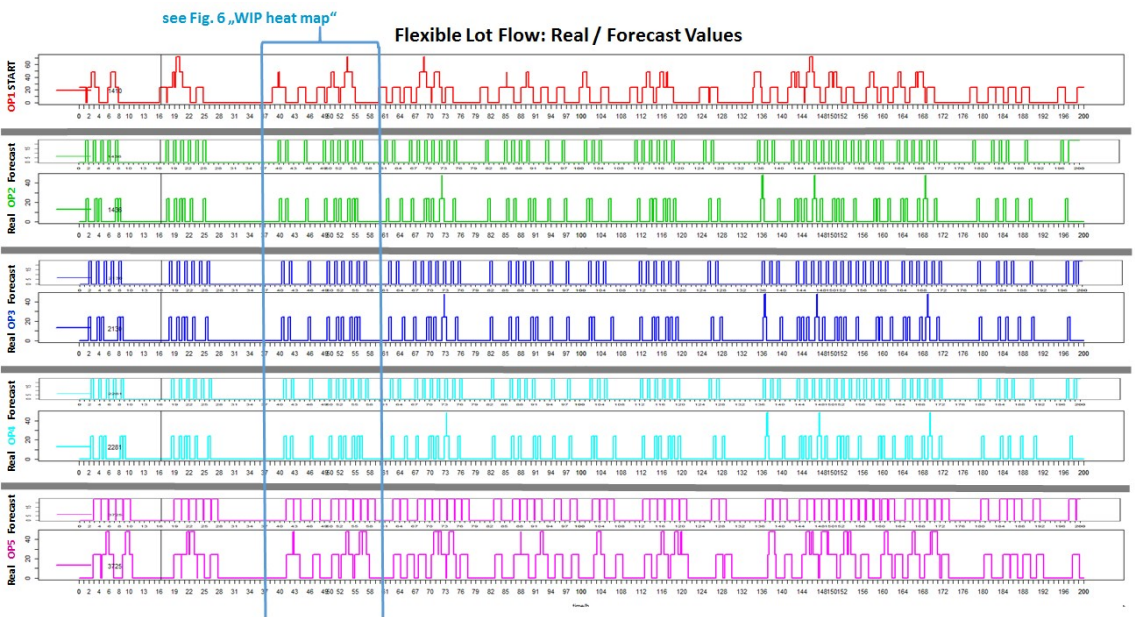


Figure 5: Comparison real values / forecasted values; flexible lot flow

	02.01.2017												03.01.2017											
	02.01.2017												02.01.2017											
	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00
1410				11,93	30,82	24,00	4,38		6,39	24,00	5,20		12,09	23,48	14,23	37,93	18,97	22,27	50,37	35,21	0,94			4,75
1436				0,37	11,00	11,43				11,33			0,52	10,85	2,67	13,71	6,33	10,31	12,39	11,35				
2130					11,29	8,19	3,08			7,47	3,81			11,28		11,29	11,24		12,69	21,07				
2281					1,71	9,57	11,25				11,23			1,87	9,41	4,03	13,63	4,83	11,24	11,73	10,78			
3725					14,43	30,11	24,00	1,37			8,96	24,00	1,87		14,59	20,24	16,79	37,23	16,32	26,91	47,80	29,94		

Figure 6: WIP Heat map: flexible lot flow; heat map time interval indicated in Figure 6

The next figure shows a calculated measure of the difference between forecasted and real data. The calculated distance measure CCD has been used (CCD measure : cross correlation distance, distance measure based on the cross correlation function).

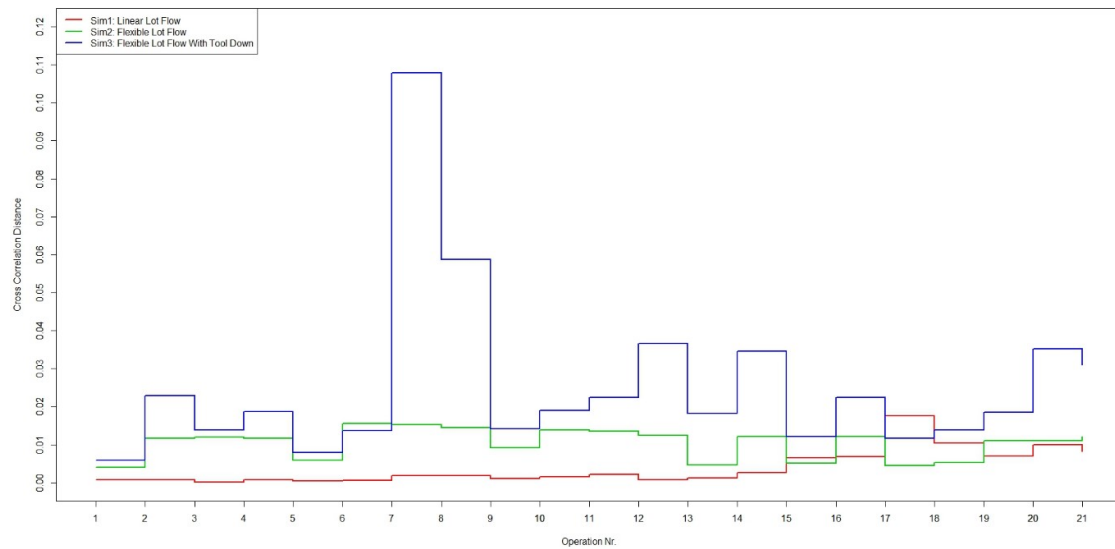


Figure 7: Comparison of the forecast quality based on calculated distance measure CCD

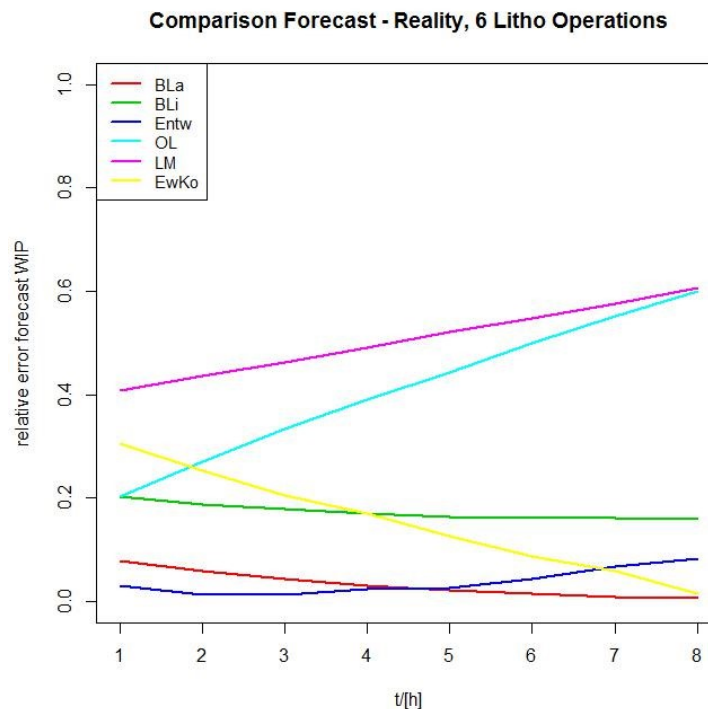


Figure 8: CCD error of forecasting; date: 2018-06-27; starting 11:00 (first point "1"), measuring 8 hours of forecast; the next point ("2") initializes a forecast at 19:00 (8 hours forecast again), and so forth.



Finally, Figure 8 shows the forecast quality using Elmos production data. Elmos produces a large amount of different products (including different technologies) and never the less using similar or same equipment. In order to test the forecast, the lithography area has been selected, covering typical operations such as coat (BLa), exposure (BLi), develop (Ent), line measure (LM), overlay (OL), develop-control (EwKo).

We may note that the behavior of the overall production system can qualitatively be studied by using different kinds of heat maps. The next figure shows the average amount of wafers during 1 hour in 1 operation (BLa etc.).

	27.06.2018													28.06.2018									
	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	00:00	01:00	02:00	03:00	04:00	05:00				
► BLa	5,27	19,89	28,69	36,07	47,81	59,59	74,15	83,88	84,12	94,52	99,39	100,00	95,28	99,36	96,52	88,57	85,09	80,16	80,41				
► BLi	3,79	21,78	28,69	33,57	31,28	33,04	32,20	36,64	38,21	36,31	37,59	39,78	43,56	51,02	53,05	56,68	60,06	62,17	66,10				
► Entw	3,91	11,91	11,42	13,71	17,24	16,24	14,78	15,02	18,48	20,42	18,25	14,59	13,22	9,31	12,58	9,90	11,30	12,05	10,92				
► LM	0,86	8,33	12,59	11,41	15,93	19,50	17,65	14,46	14,28	15,68	19,96	22,66	26,37	26,29	26,25	28,34	25,34	22,50	25,87				
► EwKo	1,30	7,52	8,45	7,12	3,14	9,76	6,68	5,61	4,56	0,88	4,31	8,17	11,92	11,66	10,41	7,46	14,77	21,67	11,81				

Figure 9: Time-series / operation Heat map of lots (WIP; normalized values; max. value=100)

Some measurement processes are showing high relative error, while the critical production process (BLi) is also showing a high and “stable” variability (relative error values around 0.2; WIP values constantly growing). This is already a better quality than achieved in the prior tests (Figure 3), because a couple of adaptations of the data model have been done. For further root cause analysis, the variability of each equipment may be analyzed in more detail.

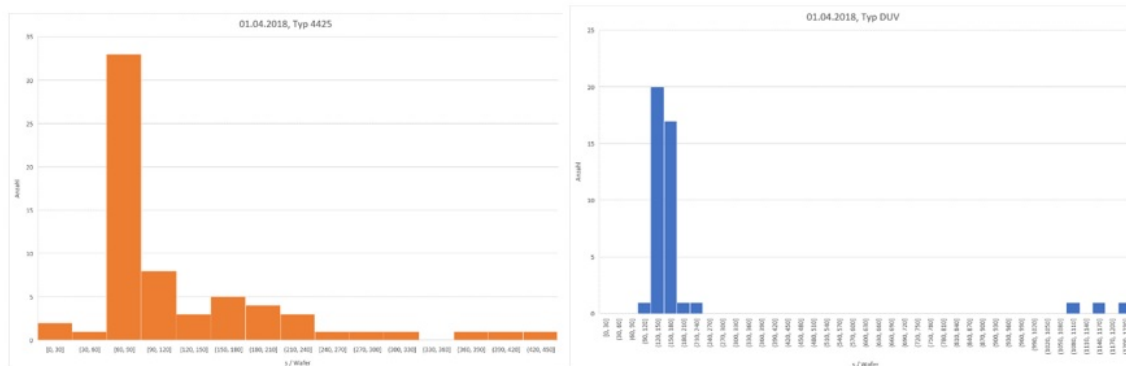


Figure 10: Detailed variability of process times of production equipment (histogram) (Datum muss noch angepasst werden)

Even more detailed analysis shows, that the distribution of the process times shows a dependency with regard to different products / product types. The analysis of the DUV histogram highlights another interesting detail (Figure 10). While usually being stable, some values show unexpected high process times. However, those values are caused by a different kind of usage of the equipment: for some lots the equipment is used as a complex cluster tool, covering coating, exposure and development at the same time. While for the most part of the lots the equipment is only used as an exposure tool.

All factors are hindering better forecast quality. However, if the analyzed reasons are taken into account, and integrated into the mapping of the data model, than the forecast quality will be raised again.

To summarize, scope of this paper is to show the capabilities of the holistic information model to offer the capability of deep and thorough analysis of the overall semiconductor manufacturing process, and offering new functionalities at the same time.

This new approach in information science demonstrates new qualities in information processing and advanced production analysis, enabling the potential for continuous optimization of semiconductor manufacturing, and within emerging globalized production networks. All companies Elmos, Infineon and SYSTEMA are working closely together, and strive to continue their successful cooperation.

## 2.19 Capturing Industry 4.0 Use Cases for Interdisciplinary Implementation

Author: Manfred Rosenberger; Virtual Vehicle Research Center Author: Alexander Stocker; Virtual Vehicle Research Center Author: Andreas Felsberger; University of Klagenfurt

*Industry 4.0 is commonly used as a collective term for various technologies, methods and applications related to the improvement of production and support processes. The introduction of industry 4.0 technologies is complex and challenging, because implementation projects involve experts from a wide variety of disciplines. This often leads to communication problems between these experts, which can endanger the success of industry 4.0 implementation projects in the long term. Therefore, suitable and above all practicable approaches are required to support the implementation of Industry 4.0 projects, so that a common understanding of the initial situation, the challenges, the target situation, the solution approach, and the foreseeable and measurable benefits is already created at a very early stage in implementation projects. Motivated by articles in the field of requirements engineering and case study research, this paper presents a lightweight approach to support the implementation of industry-40 technologies by assisting interdisciplinary project teams. This approach foresees a structured description of the industry 4.0 use case as a result of an interaction process between a dedicated requirement manager and an industrial use case owner, abandoning specialized and technical vocabulary, which can be understood by persons from different domains. The approach presented in this paper was developed and successfully applied in the industry 4.0 research project SemI40 to support capturing 30 different industry 4.0 use cases.*

### Industry 4.0 – Complex Projects and Challenges

Since the term “industry 4.0” (Kagermann et al., 2013) has been established in the German-speaking world, the so-called “informatization” or “digitization” of production processes or entire value chains has been on the agenda of many event formats. Regardless of this, the term industry 4.0 is still used very vaguely. The discussion about the increasing digitization of production and business processes brings back memories of discourses already held towards the end of the last millennium. Even then, both scientists and practitioners used the term “eBusiness” to describe precisely phenomena that are now being associated with “digitization” or “digital transformation”. Already more than two decades ago, research focused on how computer and internet technologies can contribute to transforming both business processes and business models. In the end, many discussions ended in the dot-com bubble. While eBusiness focused on business processes and business models, the Cyber-Physical-Systems (CPS) paradigm aimed at the physical product that continuously transforms from a mechanic and a mechatronic to a cyber-physical system by using Internet of things technologies (Zhou et al., 2016; Denger et al., 2014). Any production unit can be analogously described as a cyber-physical production system (CPPS) established by using industrial internet technologies. All these terms have in common the desire to solve operational challenges by adopting modern information and communication technologies. However, in many cases the organizational context in which these technologies are to be introduced is often not considered well enough to reap the full benefits of the technology, as such projects are driven by technology-focused people.

Although industry 4.0 implies the merge of information and communication technologies with production and process technologies, it is still challenging for people independent of their particular domain to describe concrete application cases (Richter et al., 2015). This holds especially for industry 4.0, which has an extremely broad spectrum of fields of applications (Hermann et al., 2015) with respect to the transformation potentials offered by technology. Such a practice of technologies affecting organizations is well known from the past, e.g. through “Web 2.0” or “Enterprise 2.0” (Stocker und Tochtermann, 2012, Stocker et al. 2012, Stocker and Mueller, 2016), which aimed at bringing web-technologies and applications into organizations. While approaches including

Industry 4.0, CPPS, Web 2.0, and Enterprise 2.0 often imply the existence of technology-centered or technically dominated phenomena, since projects in this area generally aim at the use of information and communication technologies within or between companies, social and economic aspects and consequences are very often neglected in the discussion during the introduction of such solutions (Stocker et al., 2014).

But Industry 4.0 must not be understood as a purely IT-related topic, because its introduction influences (and is influenced by) work organization and work design (BMW, 2014, Hannola et al. 2018). It also has an impact on knowledge sharing, learning and the skills required of employees working in increasingly mechanized environments (Stocker et al., 2014). For example, current information and communication technologies such as data glasses can be used to support employees in manual activities or better guide them through these activities (Brandl et al., 2014; Stocker et al., 2016). Thanks to the possibilities of comprehensive data analysis and steadily increased computing power, new algorithms can recognize patterns in process and quality data in ever shorter times in order to detect or forecast quality-relevant events (Stanisavljevic and Spitzer, 2016) and provide decision support or even decision making. In the medium to long term, this may lead to a change in the work of quality managers. It can be assumed that the support of human decision-making processes through algorithms will lead to comprehensive changes in human work in the long term (Campatelli et al., 2016). The switch from classic maintenance and service processes with fixed intervals to predictive maintenance planning made possible by the observation and analysis of sensor data could furthermore revolutionize the complete work design of service technicians in the future (Lee et al., 2014). In the future, modern worker assistance systems will ensure even more strongly that time-consuming and strenuous manual activities are carried out or supported by machines in order to give people more time for knowledge-intensive or innovative activities.

Successfully implementing technologies into organizational environments is a challenging task. Many projects fail in providing the envisaged benefits. Hence, the aim of this paper is to provide a lightweight methodology supporting interdisciplinary implementation teams in the execution of industry 4.0 projects which has been developed (and was executed) within the SemI40 project. It has been successfully applied in 30 industrial use cases contributing to goals such as balancing system security and production flexibility, increasing information transparency between fields and enterprise resource planning (ERP), managing critical knowledge decision making and maintenance, achieving fab digitalization and virtualization incl. simulation, and establishing automation systems for flexible distributed production.

In general, all industrial partners of SemI40 had a relevant challenge in their fab to be tackled in SemI40. To solve this industrial challenge, a technical solution approach was developed in close cooperation between industry partners, technical partners, and scientific partners from different domains within the SemI40 consortium. An impact was supposed to be generated through implementing the technical solution approach in the fab of the industry partner (who is facing the challenge) within an industrial pilot. The impact was generated as a result of a close cooperation between industrial partners, technology partners and scientific partners working well together in interdisciplinary implementation teams. The presented approach of capturing use cases for interdisciplinary implementation teams has contributed to the successful implementation of the SemI40 use cases – and thus to the success of the project.

### **Requirements, Use Cases and Case Studies**

The presented approach was motivated by several publications in the fields of requirements engineering and case study research. In order to successfully carry out research and development projects, requirements for the (socio-technical) system to be developed should obviously be well-known at the beginning of the project. Hence, requirements engineering as a systematic approach to specification and management of requirements deals with one grand challenge of developing systems: It

wants to know, document and manage all relevant requirements of all relevant stakeholders and then build consensus among stakeholders about these requirements (Pohl und Rupp, 2011). As a rule of practice, requirements are commonly described in long lists at a high level of technical detail following a certain scheme involving many experts from different domains. However, there is a danger that these people involved in the requirement elicitation process will delve into a level of technical detail way too early and then get bogged down without having gained a sufficient understanding of the big picture beforehand.

Use case diagrams are well known in software development projects and often used in connection with the collection of requirements and applied in the communication of requirements to stakeholders involved. The term "Use Case" was introduced by Jacobson et al. (1992) to describe the functionality of a technical system to be developed based on simple graphical models. In general EU-project language, however, the term "use case" is often used at a very high level, for example to describe a technical solution approach in textual form to communicate it to stakeholders. The term use case is moreover also often used as a synonym for "user story" or "user scenario" to document the application of a system to be developed as prose text in an application scenario.

The term "case study" is also used very differently today, too. According to Yin (1993), a case study is an empirical investigation that examines a contemporary system in its real-world context. According to Gerring (2004), a case study is an intensive examination of a single unit with the aim of understanding a larger whole. In summary, a case study usually examines and describes complex systems that are difficult to distinguish from their context. However, the specific structure in case studies to describe a case can help to better understand the big picture of a technical system implemented into an organization. Hence, much from case study research can be applied to gain a better understanding of the big picture.

Senger and Österle (2004) present requirements derived from business engineering for a generic case structure (for case studies) by highlighting three essential core areas in the transformation of traditional business solutions to new business models of the information age by the help of technology: The old solution (as-is situation) and the resulting challenge for the organization are described on three levels. Strategy, process and system is the first core area. The transformation project with its initiators and project goals as well as the project process and the critical success factors perceived during implementation represent the second core area. The third core area is the new solution (of the information age), which is described at all levels of business engineering along with a discussion of cost and benefit aspects.

### **Results: An Approach to Capture Industry 4.0 Use Cases for Interdisciplinary Project Teams**

In practice, the introduction of industry 4.0 technologies in organizations is a complex socio-technical project, because such implementation projects must involve stakeholders from different disciplines, from process engineering to mechanical engineering, electrical engineering, chemistry, statistics and computer science – to name a few – and initialize a change in work. Usually, misunderstandings often arise between participating experts from different domains, especially in the early phases of projects (and some of them may even be never fully clarified). Although a similar technical vocabulary is used, the terms may and - in most cases - will be interpreted in a different way by different people leading to communication barriers. Such misunderstandings can obviously endanger the success of industry 4.0 projects already at an early stage, and therefore must be prevented as far as possible (Rosenberger and Stocker, 2017).

The approach presented in this article aims at a solution to solve this communication dilemma. It differs two important key players responsible the implementation of an Industry 4.0 project into an organization, the requirements engineer (who may come from a very different discipline) and the technical project manager (as use case owner) who manages the implementation project. The

requirements engineer should obviously actively support the technical project manager as use case owner already in the early project phase or even in the pre-project phase by asking the use case owner specific questions (why, where, what, how, ...) about the project to develop a description of the industry 4.0 project, which is understandable by people from all involved domains. Such a capture of the big picture should avoid technical vocabulary as far as possible and should be understood by all persons involved in the implementation team.

Essentially, the term "use case", which is frequently used in the development or introduction of information systems, is used as a synonym for the term "application scenario". A use case or application scenario is a "boundary object" (Star and Griesemer, 1989), to break the boundaries between the different disciplines involved and to enable a common understanding among the implementation team about the core facts of the industry 4.0 introduction project. As a use case owner, the project manager is responsible for defining and communicating an application scenario to all stakeholders on the one hand and benefits from a structured use case description written in simple words on the other. The developed procedure included five steps to be executed idealistically:

- A dedicated requirements engineer describes the benefits of a systematic use case engineering approach and then defines the most important terms.
- A requirements engineer establishes (and explains) the structured template for the description of the industry 4.0 deployment project and provides it to the use case owner to capture the most relevant aspects.
- The requirements engineer analyzes the completed use case template on draft level and then asks additional questions to better capture the big picture of the use case. This especially involves a series of where, what, why, and how questions.
- Requirements engineer and use case owner (together) create a first description of the industry 4.0 use case following the defined structure. This should be an interactive process, while each iteration increases the quality of the use case description.
- Both iterate the use case description, involving additional stakeholders of the implementation team to validate the content of the description and to make sure that the content is understood well by all. They then freeze a version, satisfies the needs of all people involved and is able to communicate the big picture of the use case.

Since the application of such a procedure naturally involves a lot of effort, the first step is to convince the person responsible for the industry 4.0 implementation project of the benefits of a systematic use case development approach. To this end, the procedure provides for the introduction of the two terms "use case" and "case study" for the respective project context in order to create a common understanding of the application process among all project participants. It is essential that this definition of the term as part of the project glossary already shows how use cases (and case studies) are structured in terms of content in order to promote a common understanding of the industry 4.0 implementation project.

In a second step, a structured template for the industry 4.0 implementation project is set up to support the project manager as the industry 4.0 use case owner in the documentation and communication of the industry 4.0 use case to be implemented. Based on Senger and Österle (2004), this template should be divided into three sub-areas: initial situation and challenge, target situation and solution approach, as well as the expected effects of the use case implementation. The task of the use case owner is now to (try to) fill this template (on draft level) in order to further sharpen his own understanding of the use case. For this purpose, he may already be supported through a set of questions on the contents of the use case description. Filling out this template already helps the use case owner to better communicate with the experts from the various domains within the implementation team who, as project members, are jointly responsible for the success of the industry



4.0 implementation project.

<p>A <b>Use Case</b> is an in-advance documentation of (1) a business-relevant challenge or problem (2) of an industrial partner - 'Enduser' (3) concerning a specific manufacturing process or a specific set of manufacturing processes and/or support processes (4) and briefly describes the envisaged technical solution approach, (5) including the expected impacts (benefits) on a technical, social, and business level.</p>	<p>A <b>Case Study</b> is an ex-post documentation of (1) how an industrial partner has solved a business-relevant challenge, and (2) through which particular technical solution, (3) in which specific manufacturing process or set of manufacturing processes/support processes, (4) including the generated impacts (benefits) on a technical, social and business level, as well as (5) the success factors of the use case implementation.</p>
---	--

Figure 1: Use cases and case studies

Section	Content
Abstract	The use case is described as a management summary structured according to the initial situation, solution approach and expected benefit.
Company Description	A short company description should make it easier for the internal and external project participants to think about the use case.
Current situation (as-is situation) and challenge	The description of the initial situation and the pressure of suffering in a factory is fundamental to sharpening the objective and to defining the solution approach. The initial situation is not the same as the challenge. Both should be described using present tense.
Target Situation (to-be situation) and expected technical impact	The combination of the description of the objective and the expected benefits of the new technical solution for solving the challenge allows the definition of a measurable benefit. The target situation is not the solution approach. It should be described in future tense.
Solution Approach – enabling technologies and processes	The description of the solution approach on a technology or process level is the technical core of an industry 4.0 implementation project. Both technical and non-technical elements of the solution approach should be described.
Expected socio-economic impact	Industry 4.0 has a socio-technical nature: The expected effects of the solution approach must also be described with respect to human and organizational levels.
Knowledge and Partners	Industry 4.0 projects are usually interdisciplinary. This section describes the required competencies and partners for the implementation project.

Figure 1: The structure of a use case documentation for industry 4.0 implementation projects

In a third step, the requirements manager conducts an interview-supported analysis of the use case. It can be advantageous if the requirements manager comes a different domain than the use



case owner, as this enables the best possible identification and elimination of ambiguities in the documentation. By asking specific questions (why, what, where, how) to the industry 4.0-use case owner, the requirements manager supports the creation of a documentation of the industry 4.0-use case that is understandable beyond the involved technical disciplines. These questions help the persons involved in the development of the big picture. By filling in the structured template from the second step, the use case owner can already prepare himself for possible questions from the requirements manager.

In a fourth step, a use case draft is created, either by the use case owner or by the requirements manager. There are two possible variants: The first variant consists of transferring the entire documentation task to the industry 4.0 owner and leaving only quality assurance to the requirements manager. This means that the requirements manager will probably have to deal less intensively with the respective specialist domain. The disadvantage of the application can consist, however, in the fact that the use case description captures the view of a specialized domain. In the second variant provides the requirements manager is responsible for the use case description. This procedure forces the requirements manager to become more familiar with the respective domain and requires some corresponding technical competencies. The advantage compared to the first variant can be a more understandable use case description.

In a fifth step, a coordination process initiated by the requirements manager takes place, which is only completed when all participants of the implementation team are satisfied with the quality of the description of the industry 4.0 use case. The resulting document is finally reviewed either by the requirements manager or by the use case owner, whereby ambiguities are marked, or questions or comments are added. The coordination process leads via different revisions of the use case draft to a final description of the industry 4.0 use case.

### **Summary and Lessons Learned**

The approach presented in the article was developed in the industry 4.0 research project Semi40 and then successfully used for the systematic collection, detailing and documentation of 30 different industry 4.0 use cases, e.g. for engineering production (Stanisavljevic et al, 2018).

Experiences made have shown a lot of persuasive work to be done, especially at the beginning of a project, in order to convince all participants of the benefits of applying such a structured approach. This is understandable, as the systematic description of an industry 4.0 use case in a case structure involves a great deal of effort. Managers responsible for the introduction of industry 4.0 technologies are generally experts in their respective domains. While it was comparatively hands-on for them to describe an envisaged technical solution approach, the linguistically simple formulation of the initial situation, the challenge and the expected benefits proved to be more difficult.

In the further course of the application of the approach, it turned out that in most cases it was perceived as more useful if the requirements manager took over the responsibility for the use case documentation, while the use case owner remained responsible for the review of the document. On average, two to three iterations of a use case draft were necessary until the participants of the implementation team were satisfied with the result of the documentation. Above all, the formulation of the expected benefits at the human and organizational levels posed challenges for those involved, since taking a socio-technical system view of industry 4.0 is complex on the one hand and often unusual for technicians on the other. In many cases, the long-term effects of industry 4.0 technologies on these two design levels are not clear enough to be expressed in words. It has been shown that the requirements manager can also make a major contribution to formulating the expected benefits at the human, organizational and technological levels through asking relevant questions.

Finally, the application of such a structure approach to capture use cases in order to support interdisciplinary project teams has proved to be very useful in the project.

Semi40 - Use Cases Refined and documented in WP5			
WP1			
Secure Cyber Physical Productions Systems			
UC1.1	Production Plant Availability	IFKL	Achieving a structured documentation of the actual production network infrastructure by applying log file analysis on end devices.
UC1.2	Network Architecture	IFKL	Redesigning the production network infrastructure based on existing documentation and related internal requirements.
UC1.3	Secure Communication for Supply Chain Partner	ELMOS	Implement selected communication scenarios between supplier and customer to enable the identification of organizational and technical requirements for a secure communication across the partners in the supply chain.
UC1.4	Security Aware Industrial Network Simulation	AIT	Develop attack scenarios to simulate external attacks on a production network and evaluate security measures.
UC1.5	Enable Secured Access to Legacy Systems via Flexible Secured End-To-End Communication	IFAT	Enable network access to legacy systems in production lines using secure elements and/or trusted platform modules (TPM).
UC1.6	Anomaly Detection for Industrial Networks	IFKL	Collected typical communication and operation behaviour of end nodes in the production network is used to detect abnormal behaviour due to deviations.
UC1.7	Semantic Integration of Legacy Equipment	IFAK	Integration of conventional automatization and control systems within Industry 4.0 networks.
WP3			
Machine Learning & Automated Decision Making			
UC3.1	Health Factor for Process Patterns	IFAT	Explore relations between patterns in the wafer test data and their relation to single production steps by applying novel data transformations to wafer test data to make process patterns visible and classify them in a second step.
UC3.2	Traceability and Automated Decision Making for the Photo and Copper Process - with Focus on Embedding	AT&S	Improve the data collection process in the PCB production and perform a correlation analysis with quality data for a better control of the production process.
UC3.3	Optimized End-of-Line Test procedure for Combustion Engines	AVL	Identify significant correlations in data from an engine development and production project to describe engine characteristics and enable an optimized and modular engine testing procedure.
UC3.4	Pilot Line for Advanced Process Control	AMKOR (NANIUM)	Develop new methods of data analysis to identify unknown correlations between single process parameters and their influence on the contrast deviation of the investigated routings.
UC3.5	Advanced Process Control for Pulsion Tools	IBS	Implement a new way of configuring and controlling the plasma doping procedure to result in a more stable and easier to configure doping process.
UC3.6	Advanced Dispatcher & Scheduler	IFAG	Implementation of an automated reticle dispatcher & scheduler for the lithography cells of a production environment.
WP2			
Agile Facility			
UC2.1	Preventive Maintenance Scheduling in Cooperation with Fab Workload and Workload Simulation	BOSCH	Visualize optimal maintenance time for individual equipment and predict plant utilization through simulations thereby considering "on time delivery".
UC2.2	Advanced Dispatch Control	IFAG	Analyze the existing planning of the production volume and the machine assignment as well as current events in the supply chain and automatically propose an optimized corrective action.
UC2.3	Demand monitoring via SPC	IFAG	Characterizing the ordering behaviour of a given customer with descriptive statistics, define threshold values and provide a control mechanism to adjust production requests.
UC2.4	Inkless Chiptracing History	SEMIKRON	Replace optical marking of chips by a continuous data based solution (e.g. a digital wafer map) and adjust the production process to ensure a complete tracking of individual power electronics modules.
UC2.5	Production of Refractory Metal Sputtering Targets	PLANSEE	Develop new control and optimization tools for production steps by extending established methods with tailored mathematical process models.
UC2.6	Material Tracking for Non-Product Wafers	ELMOS	Implement the administration, planning, controlling, and monitoring of NPWs into the manufacturing execution system.
UC2.7	Sub-Equipment Monitoring	IFAG	Implement additional sensors to measure additional process characteristics and take them into account in a virtual process model to support the process control.
UC2.8	Cycle Time Prediction	AMKOR (NANIUM)	Development of flexible planning models based on all planning-relevant information concerning production orders to be captured, which are capable of validating different production equipment utilization scenarios.
UC2.9	Process Simulation and Measurements Data Integration for Epitaxial Reactor Automation	LPE	Develop and validate an accurate mathematical model of the epitaxial reactor process to predict the next required refurbishment for optimized maintenance.
UC2.10	Bottleneck Detection	ELMOS	Develop an analysis and forecasting tool to consider the expected capacity utilization and adjust the production and maintenance accordingly.
WP4			
Virtualisation / Digitalisation			
UC4.1	Fleet Management for Self-Navigating Transportation Systems	IFD BOSCH	Implement mobile and autonomous vehicles to load and unload single manufacturing tools with production lots within a demonstrator environment for a semiconductor production and design the interaction of these vehicles with the MES.
UC4.2	Hybrid Transportation Systems	IFD	Establish a flexible hybrid transportation system to be controlled by the MES, combining elevator systems, rail-guided systems as well as autonomous mobile transportation robots.
UC4.3	Online Tool for Real-Time Visualisation of Fab KPIs	IFD	Capture all data relevant for the production and make it available in a high-performance IT system to enable a real-time visualization of KPIs, allowing drilldown of process flows, product groups and production lots.
UC4.4	Image Recognition within the Production Line for Quality Assurance	IFD IFAT	Exploit the potential of image processing techniques to detect defects on the macroscopic level after the wafer processing has been completed.
UC4.5	Energy Management	IFD IFAT ELMOS	Develop a high-performance monitoring approach to permanently monitor the power consumption of complex machines and supply systems.
UC4.6	Assisted Area Simulation	IFAT	A simulation is addressed to optimize production utilisation and transportation paths. Therefore, transportation paths will be visualized on displays mounted on trolleys which are updated frequently.
UC4.7	Planning of a Transport System Extension	IFAT IFD	Remove an existing bottleneck in the transport system, which connects various facility levels and optimize the placement of an additional floor-to-floor elevator by simulation.

Table 3: Project Use Cases

In summary, the procedure involved the following steps and roles:

Step	Person in charge	Task	Artefact
1	Requirements manager	Describing the benefits and scope of the approach and defining terms	Value proposition, glossary
2	Requirements manager	Development of a structured template for the industry 4.0 implementation project	Use case template
	Use Case Owner	Reflecting the use case and initial filling out the template	Filled use case template (draft)
3	Requirements manager	Analysis of the use case and conduct interviews or workshops with the use case owner	Protocols of conducted interviews and workshops
4	Requirements manager or use case owner	Creation of an initial description of the use case following the defined structure	Use case Draft
5	Requirements manager, project implementation team	Iteration of the use case description with all participants of the implementation team. Elaboration of the content in the use case template.	Final and mature use case description

Figure 1: Approach to support the introduction of industry 4.0 technologies

## 2.20 Exploiting Advances in Video Technology for Use in Smart Manufacturing

Author: Rene Kaiser; Know-Center research center Graz

*Video content and technology is an integral part of our private and professional lives. We consume news and entertainment content, and besides communication and learning there are many more significant application areas. One area, however, where video content and technology is not (yet) utilized and exploited to a large extent are production environments in factories of the producing industries like the semiconductor and electronic components and systems (ECS) industries. This article outlines some of the opportunities and challenges towards better exploitation of video content and technology in such contexts. An understanding of the current situation is the basis for future socio-technical interventions where video technology may be integrated in work processes within factories.*

### Disadvantages, Constraints and Risks

Certainly, there are downsides to utilizing video technology in factory environments as well. Investing in technology obviously entails cost for procurement, integration and maintenance of infrastructure and devices. Transmitting and recording significant amounts of video data requires IT infrastructure, bandwidth, storage capabilities as well as dedicated interfaces to retrieve, analyse and engage with content.

Apart from infrastructural aspects, another concern is the acceptance of the dedicated users who might have some freedom to decide to which degree to make use of video tools. Any camera that doesn't capture only machines but also monitors employees conducting their work on a permanent basis can become a serious issue with respect to privacy, employee rights and labour law. Even in videoconferencing where an employee takes the conscious decision to start and end the process of being captured by a camera, several participants in our interviews reported that they feel rather uncomfortable and insecure in such situations. Specifically, some who had experience using telepresence tools in meeting rooms felt disturbed seeing themselves when the interface included a self-view. It is yet to be evaluated to which extent such concerns mentioned in the realm of office spaces apply as well for remote video communication within the factory environment.

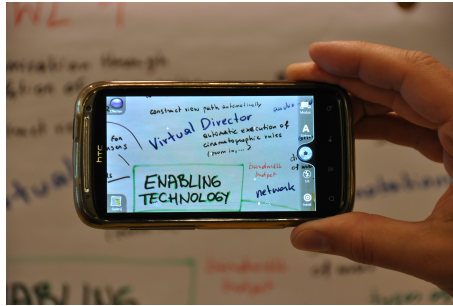
For companies themselves, any remote live video connection bears a risk for the protection of their sensitive, confidential intellectual property. Hence, many companies decide not to allow remote video streams from their factory at all. As a downside, minimizing this risk rules out any benefits based on remote collaboration.

### Opportunities and Example Application Areas

In smart factories of the semiconductor and ECS industries complex and multifaceted work processes are taking place. They naturally vary from company to company. Nevertheless, there are common areas which can be talked about in general terms, and in which video content and technology could be employed to provide added value. The following lists some of these relevant areas and concisely discusses video technology which could be utilized to a greater extent.

*Video communication* or *telepresence* is a basic feature which can be part of many services and work processes and several production companies already use, for internal communication as well as with external sources, for remote maintenance, collaboration with partners in innovation endeavors and co-opetitive settings, etc. The context of such applications is very different from private social video communication or business videoconferencing in office and meeting room environments.

Beyond 1:1 direct video feeds, in complex communication setups, Virtual Director approaches [104] [51] could help users to make best use of multiple concurrently available video streams. A Virtual Director is a software component which automates the decision making tasks of a human director in a TV broadcast, i.e. taking decisions what camera view to show, and when to switch to another



view. However, in order to adapt this approach for use cases in smart factory environments, first a deeper understanding of user/stakeholder needs and constraints needs to be obtained, especially regarding the user interaction with the system itself:

- What specific features and capabilities are required in specific application contexts within factories?
- What forms of user interaction are suitable to steer semi-automatic content selection and adaptation, given constraints like gloves and cleanroom clothing?
- What characteristics of a viewing experience are relevant to enable efficient and accurate work for both repetitive and challenging tasks, while maintaining work satisfaction long-term?

In environments like clean room production fabs where the number of employees on the shop floor keeps decreasing relative to the increase in productivity, and mobile robots are already present, apart from fixed and human-worn cameras, telepresence robots [218] and indoor/outdoor video drones [150] may prove to be useful, as such mobile cameras quickly allow to reach and inspect remote spaces.

For *machine monitoring* and *steering*, dashboards mostly include textual and abstract graphical representations of processes and states. Integration of video feeds could speed up cause analysis during costly breakdowns or help to visually determine from remote what sensors might not be capturing. Where video is already established in many processes without a human in the loop are machine vision applications, i.e. real-time video analysis coupled with fully automatic decision making, for example to filter out defective products or production equipment.

In some factories, simplistic Virtual Director behaviour is already deployed to enable automatic camera switching in order to help operators follow items along a linear production workflow across multiple fixed camera viewpoints.

For highly monotonous monitoring tasks where many machines have to be observed in parallel, and where fast and correct operator intervention is required in case of errors (related approach described in [103]), specific support needs to be designed and intelligent video processing may be part of novel solutions assisting operators to that end.

While *documentation* for the sake of quality insurance and traceability is typically based on (machine) data logging, in certain cases it would make sense to use video as well. To make use of large amounts of recorded footage whenever necessary, dedicated interfaces for retrieval and analysis are needed as well, but this is typically handled outside the production environment.

*Providing individual instructions* for workers is an aspect that could become considerably more important due to the Industry 4.0 trends of product customization. Optimizing for *lot size 1* leads to a tendency of less repetitive tasks in some areas and implies that workers require more information about the current products being processed. This is a challenging aspect especially when a considerable amount of products is to be handled in parallel. For ever-changing information and instructions, selective use of video content might be beneficial compared to primarily textual interfaces. More descriptive directions and conveying a better understanding of the desired output may increase unambiguousness of instructions and throughput for manual tasks. It may also help to

reduce mistakes.

For companies, using video for this purpose also depends on efficient means to capture content. While technically it may be straightforward to capture instructions as video clips, one learning of the aforementioned interviews was that many employees have profound reservations recording something not being a cinematographic expert, and especially recording themselves explaining something or showing how a task can be conducted. This is especially the case when video content is supposed to be shared with external parties such as customers, but also when recording informal instructional videos for colleagues.

Apart from concerns to capture evidence of any mistake, people tend to get overly ambitious with respect to the cinematic quality of their recordings. Rather than aiming for what is good enough for the purpose at hand, some employees seem to strive for what they are used to when consuming entertainment content in their spare time (TV, cinema), and do not want to share content that does not meet these aspirations. The fear of making oneself replaceable is a further reason affecting knowledge sharing in this form. This is one besides many other factors keeping employees from externalising and documenting knowledge, as known from knowledge management research.

*Learning or employee training* is another aspect where video content is already used in terms of courses and seminars. However, the uptake of Industry 4.0 approaches might change requirements towards more short-term, in-situ knowledge transfer and refreshment and require new kinds of video-based technology support. While one current concern to that end is the effort required to keep contents up-to-date, future intelligent video technology may be more flexible by automatic means, and capable of content adaptation and personalization with very little manual effort required.

### **Future Research**

Video content and technology in increasingly smart production areas is a topic that hasn't been addressed by research with emphasis in the past. Due to implications of the Industry 4.0 movement, however, companies keep adapting and improving production processes and so the tasks and required qualifications of workers are changing constantly as well. This development could inspire many production companies to more closely investigate how to exploit video to their advantage.

Towards designing socio-technical interventions as a combination of new tools and adapted business processes, a first step is to understand what reservations users and further stakeholders have towards innovative video technologies. Findings from other working areas such as office spaces may or may not apply to the specific context and conditions of factory environments.

A positive factor contributing towards technology acceptance may be the urgency of certain tasks, e.g. video-based remote maintenance in case of a costly machine breakdown: shyness may be overcome easier by a person understanding the urgency and consequential costs of such situations – at least this was the opinion of one participant in our interviews.

Another positive factor is the intimacy of mediated video communication, i.e. a 1:1 session with a remote person whom the local person has met before face-to-face may be much better accepted than conversations in videoconferences involving larger groups of people. Training employees, informing about what is recorded and what is not recorded, what is used for what purpose and who will have access to content etc., can be another enabling factor.

In any case, video technology needs to be used selectively. Appealing interfaces and ease of use are essential, especially when there is freedom of choice to use a tool or not [88]. Basic training on how to use video based technology, however, becomes less and less important since the young generation of workers is typically very familiar with it due to smartphone and camera use in their private lives.

From a research point of view, a key concept to explore are *experiences* in the realm of video use and consumption in factories (cf. [152]), a concept very well understood by the research community for other application domains but not for production environment contexts. But in



contrast to entertainment, for example, factory workers utilizing video content are driven by their working tasks and there is less intrinsic motivation to engage with video. How could the quality of experience (QoE) be conceptualized and assessed in fab environments?

The availability of the right knowledge in the right form can make an impact on productivity. For repetitive tasks, content presentation may help maintain concentration and not miss important cues. During the execution of very crucial tasks, assistance to focus on what is relevant can also help to concentrate, somewhat related to the concept of *immersiveness* in other domains, especially when AR glasses or VR goggles are used to interact.

Regarding the application of the Virtual Director concept, the most obvious cases we identified so far are for communication when collaborating remotely, and for operators monitoring a significant number of machines via remote video streams in parallel. Compared to entertainment services, the cinematographic appeal and dynamic visual aesthetics are assumed to play a much lesser role. The pragmatic concern of seeing what is currently most relevant appears to be more important in general terms. Nevertheless, it makes sense for production companies to investigate how to support workers in the best way possible. Small difference in tool behaviour could make a great difference especially for daily routine tasks.

One approach towards exploring possible applications of video technology is driven by thinking about its *affordances*. The concept of affordances has been described in various abstraction levels and from different viewpoints across several research disciplines, concerning both physical objects and technical tools and applications. An affordance refers to what actions a user can take with an object.

Gaver [63] defines affordances as "*properties of the world that are compatible with and relevant for people's interaction*" and stresses the importance of understanding both the needs and abilities of users, as well as the capabilities and limitations of technologies. Ideally, any video device, application or service shall serve multiple affordances, possibly even future ones not conceivable yet (extendability). Exploratory research at this stage may be less concerned with the concrete affordances of a particular tool, but rather aim to understand which affordances shall be supported overall.

### Acknowledgements

The work has been performed in the project *Power Semiconductor and Electronics Manufacturing 4.0* (Semi40), under grant agreement No 692466. The project is co-funded by grants from Austria, Germany, Italy, France, Portugal and Electronic Component Systems for European Leadership Joint Undertaking (ECSEL JU). The Know-Center is funded within the Austrian COMET Program – Competence Centers for Excellent Technologies – under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

### About Authors

**Rene Kaiser** is researcher at the Know-Center research center, Graz, Austria. His background is software engineering which he studied at FH Hagenberg. Currently he is pursuing a Ph.D. at Graz University of Technology. Rene's main research expertise and interest is in interactive access to multimedia content. He investigated real-time content adaptation and personalization and automatic execution of cinematographic principles, focusing on the concept of 'Virtual Director' technology. In the Semi40 project, Rene contributed to enabling knowledge transfer within the consortium, and approaches to balance knowledge sharing and knowledge protection. Extended bio: <https://about.me/rene.kaiser>





### **3. Roadmaps**



3.1 Road Maps - Cyber Physical Production Systems in Semiconductor Eco-System



Figure 3.1: Road map CPPS SemI40



Figure 3.2: Road map WP1 CPPS SemiI40

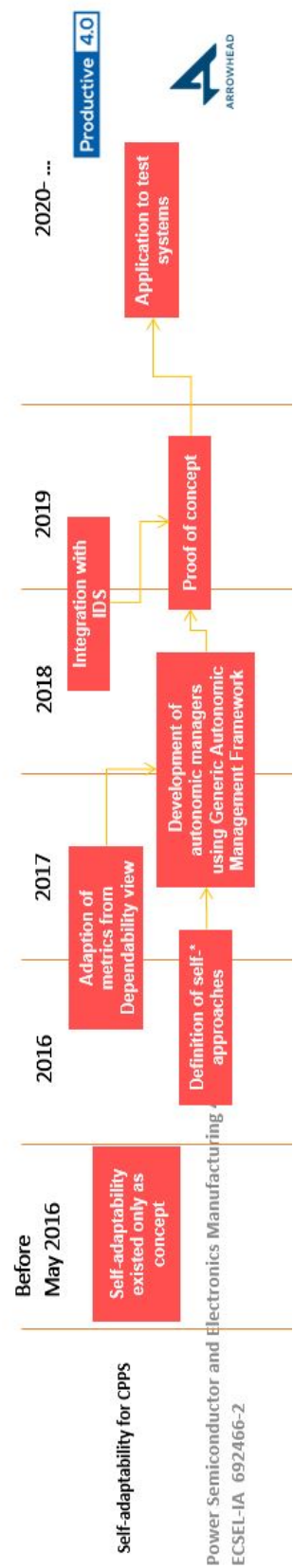


Figure 3.3: Road map WP1 CPPS SemI40



3.2 Road Maps - Agile Facility

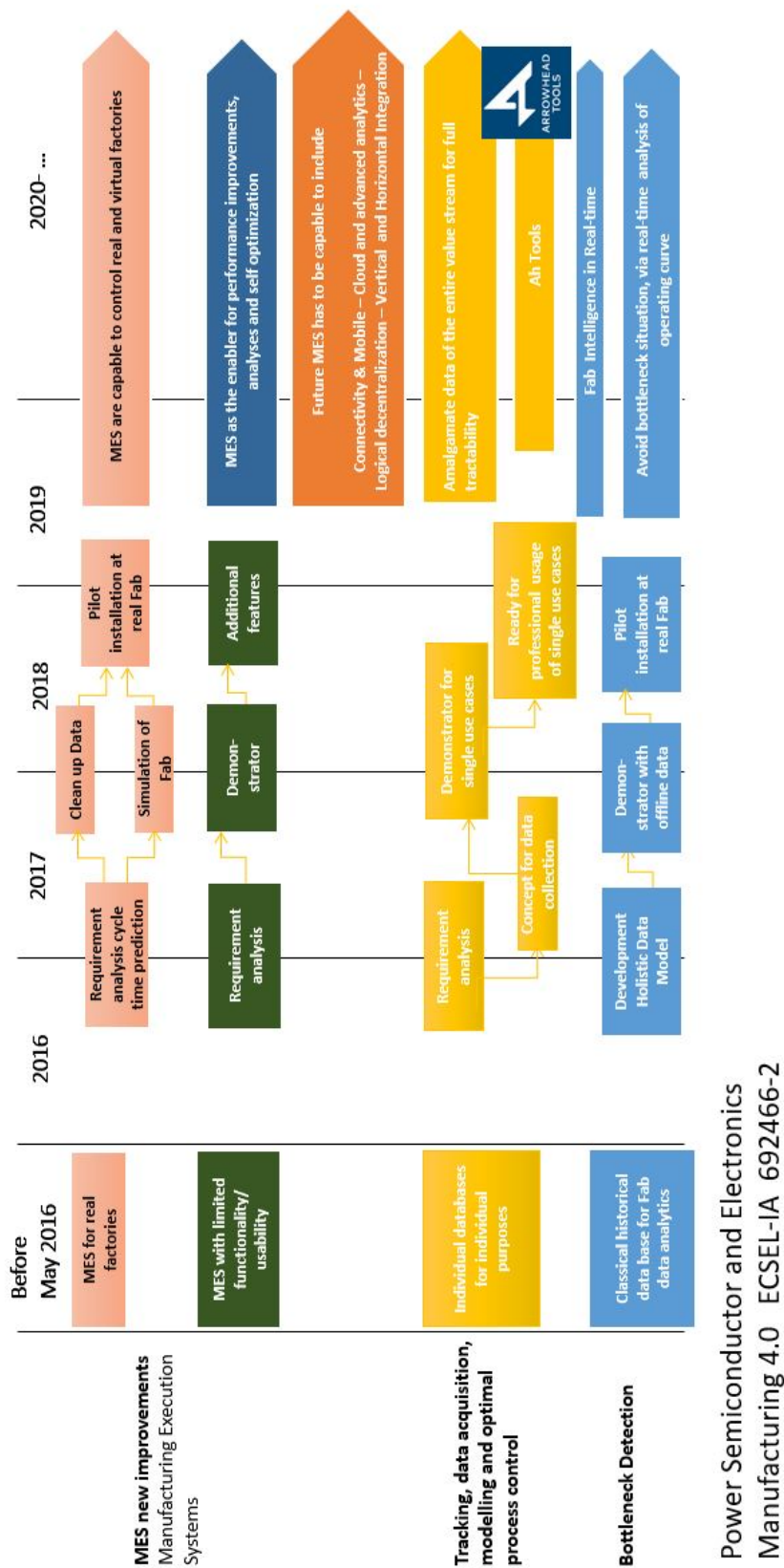


Figure 3.4: Road map WP2 Agile facility SemI40



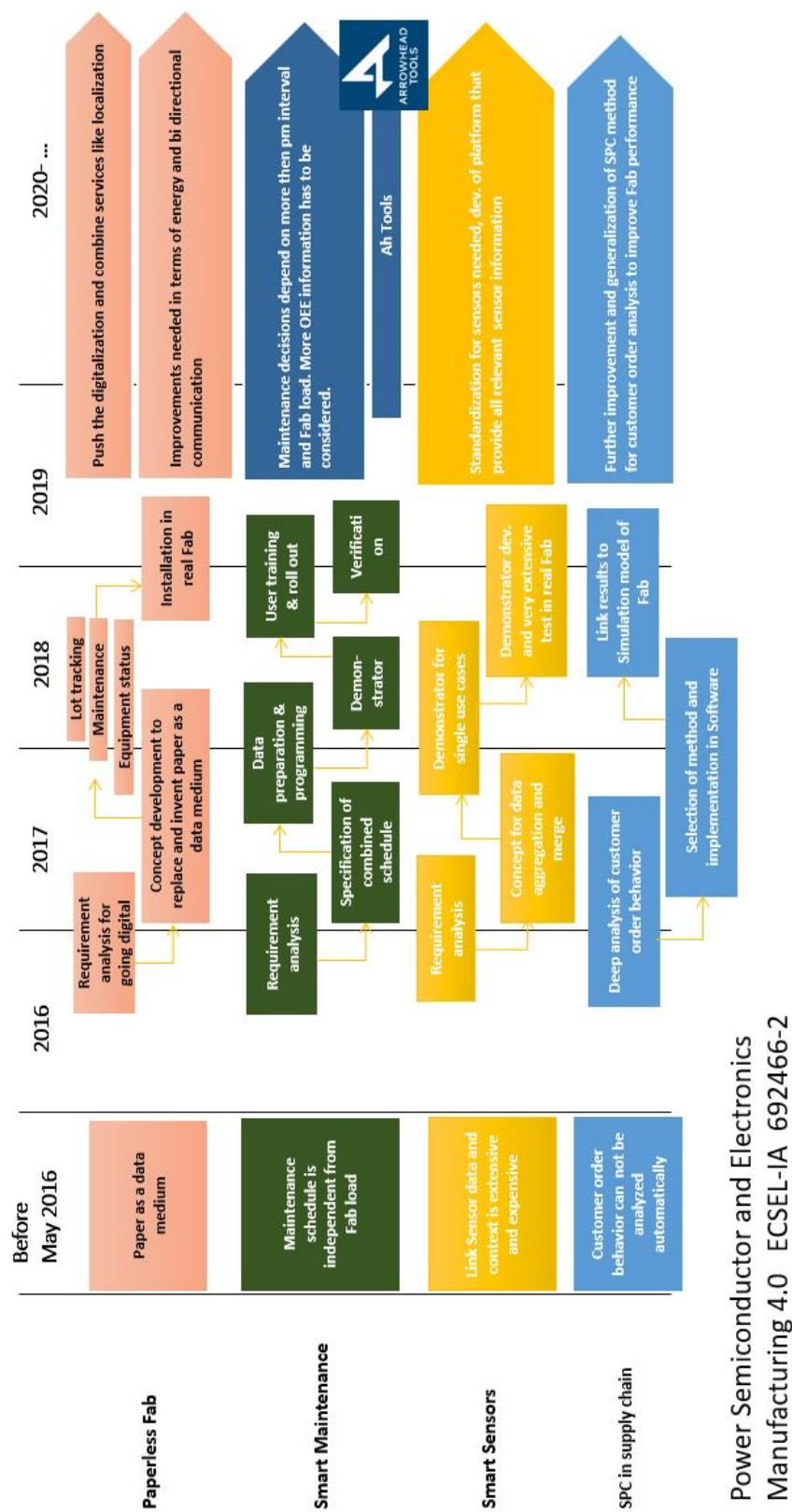


Figure 3.5: Road map WP2 Agile facility SemI40



### 3.3 Road Maps - Machine Learning and Automated Decision Making

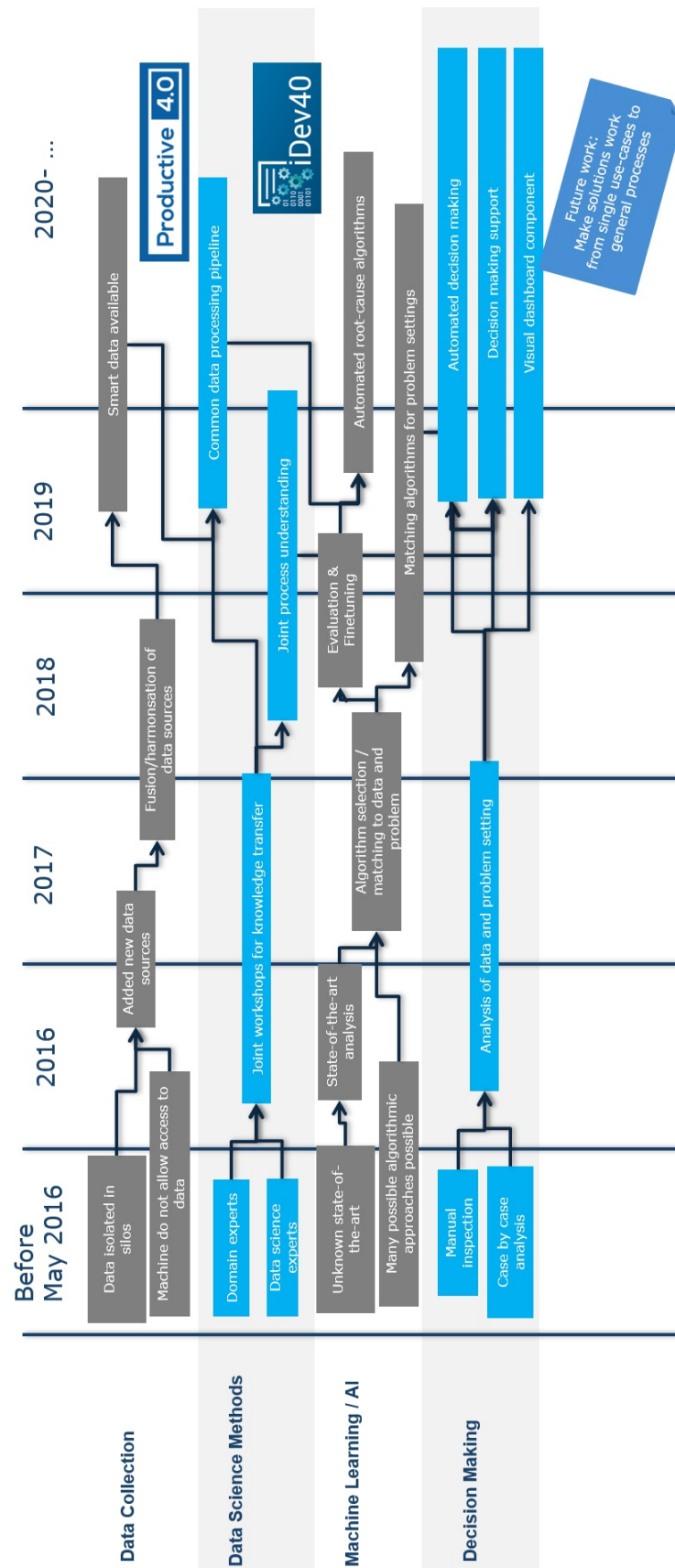


Figure 3.6: Road map WP3 Machine learning and automation decision making SemI40



### 3.4 Road Maps - Digitalization and Virtualization SemI40

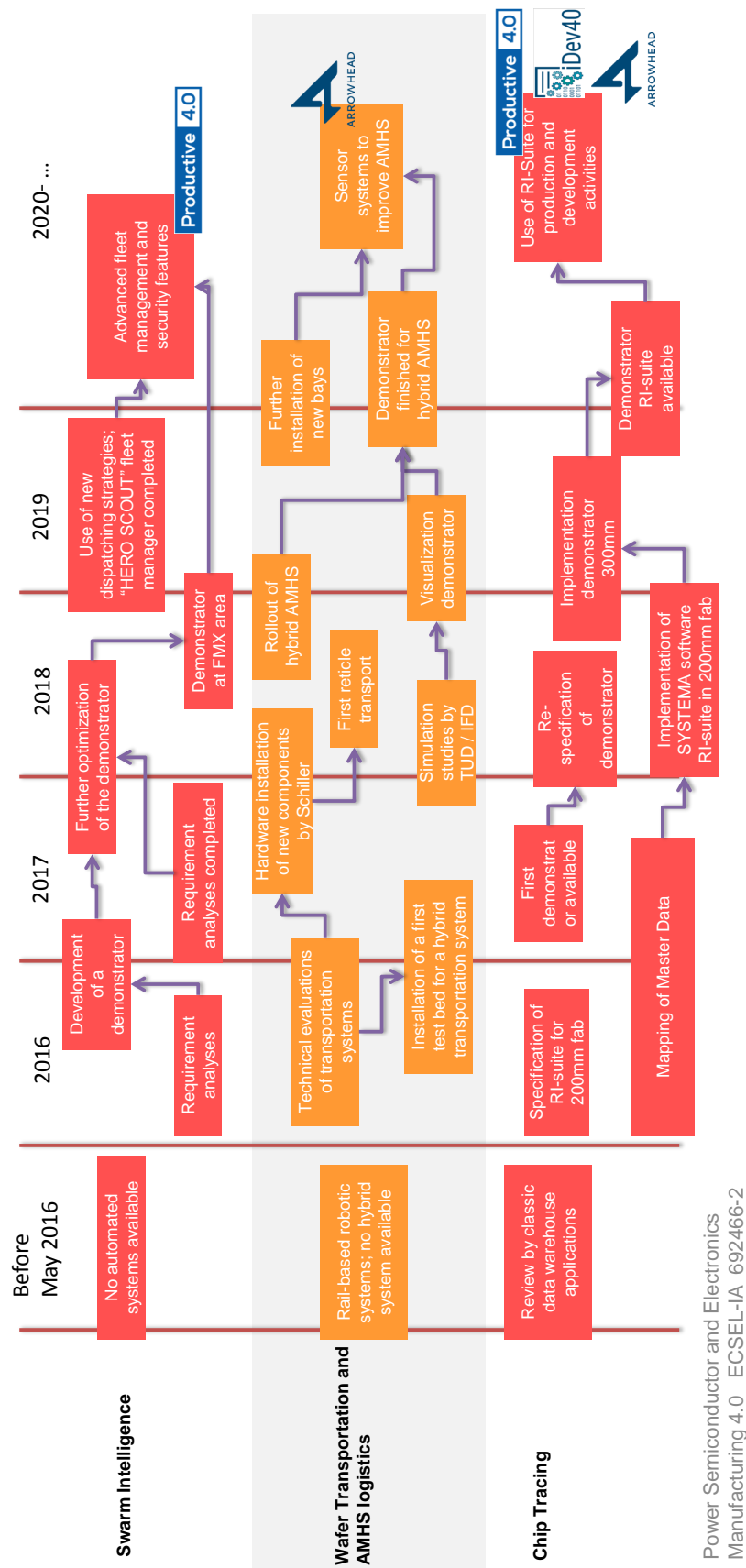


Figure 3.7: Road map WP4 digitalization and virtualization SemI40

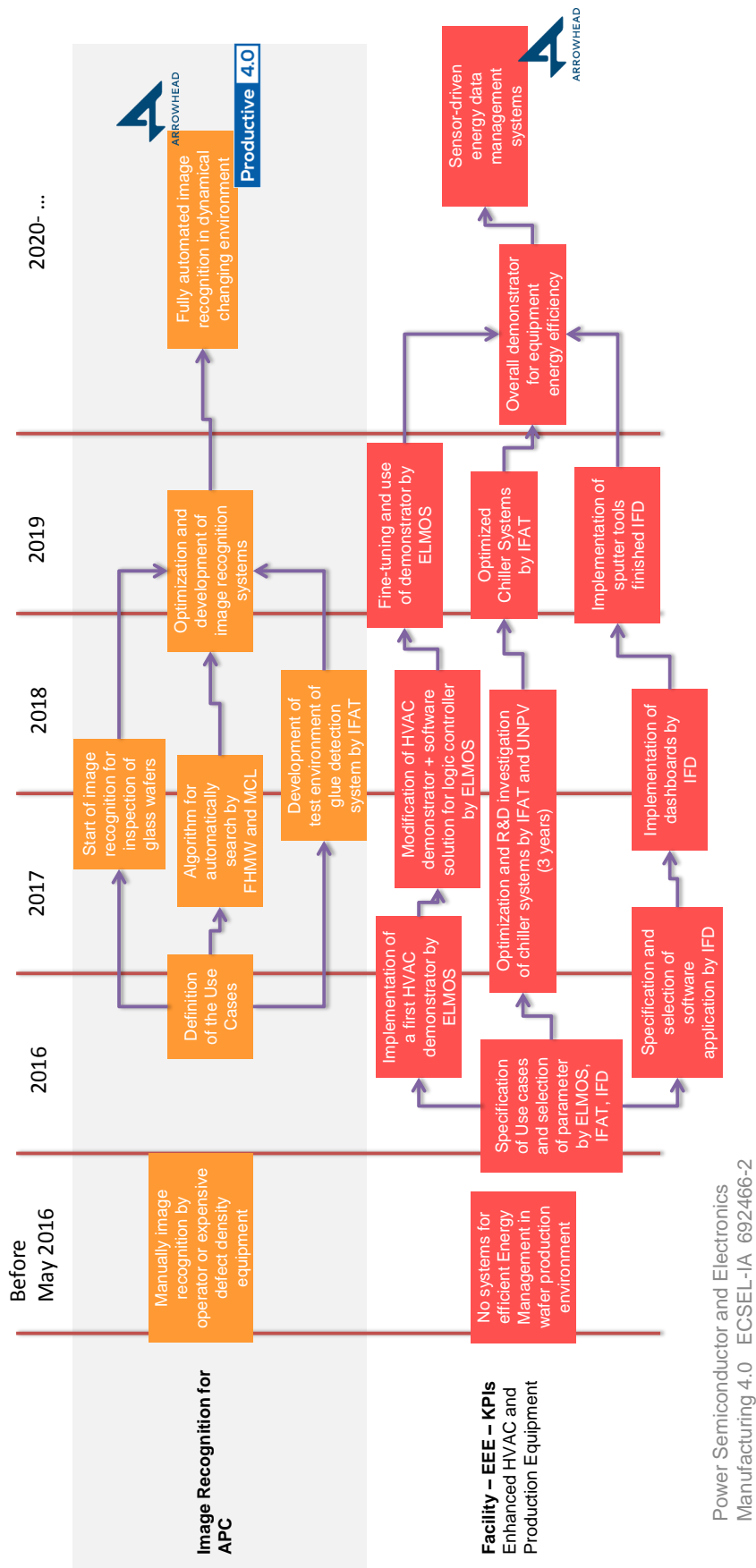


Figure 3.8: Road map WP4 digitalization and virtualization SemI40







## 4. Semi40 Research roadmap

### 4.1 Research Roadmap from a Semi40 Workshop

*Trends, Challenges and Implications for future Research Projects in the Semiconductor Industry*

Authors: Alessandro Sala and Marius Lütkemeyer; Fraunhofer Austria Research GmbH

Author: Andreas Felsberger; University of Klagenfurt

Contributors: Semi40 Consortium Partners

#### **Abstract**

*Funded RD programs on the national and international level are critical success factors to sustain the pipeline of new inventions that will fuel the economy, create jobs and deliver a competitive edge to domestic industries. Due to a constant technological development and fierce competition for global technological leadership, the semiconductor industry is particularly susceptible to fall behind if it does not stay up to date with current research. In order to take full advantage of current research, it is necessary to effectively communicate and share specific results, therefore accelerating scientific progress. While mandatory dissemination activities do already exist in research projects, there is still a lack of information shared about specific experiences and insights gained by single individuals in the project. These specific details hold huge potential to both improve methodology and content of further research. In this paper, we aim to transfer knowledge from the Semi40 project to future research activities, by presenting a semiconductor roadmap, which was elaborated in a two-day workshop. The roadmap is based on experiences gained by top industry experts during the Semi40 project, highlighting current and future challenges and trends as well as suggestions for future research projects in the semiconductor industry.*

#### **Introduction**

The exponential down-scaling of microelectronic devices, through shrinking transistor sizes and advancing fabrication processes, has led to a continuous emergence of new key applications under labels such as Smart Mobility, Smart Society and Smart Production. Enabling markets comprising of micro- and nanoelectronics, advanced materials, industrial biotechnology, photonics,

nanotechnology and advanced manufacturing systems are constantly growing, gaining increasing importance for the European industry sector (cf. Fraunhofer IMW, 2018). A comprehensive PwC analysis of the global semiconductor market from 2015 suggested that the total semiconductor market would amount to a value of US\$ 405 billion in the year 2018 (cf. PwC, 2015). In fact, posted sales in 2018 even exceeded this prediction by 15 %, reaching a total of US\$ 468.8 billion with total units shipped topping 1 trillion for the first time (cf. Rosso, 2019). Despite the constant growth of the global semiconductor market, Europe has significantly lost and is still losing ground in the volume of semiconductor-production capacities, especially compared to Asian economies. With many customers located in China, South Korea and Japan, the concern rises that the disparity between Europe and the rest of the world will increase, as the manufacturing industry tends to develop geographically close to its demand (cf. Fraunhofer IMW, 2018).

In an industry which is highly characterized by technological change, both from a product and a process point of view, RD projects are an essential success factor to prevent losing further ground. The constant drive for reducing production costs per unit and increasing the performance of micro-electronic components makes the semiconductor industry one of the most RD-intensive industries. While the European Commission is currently funding multiple large-scale research projects in the semiconductor industry, there are difficulties regarding the knowledge transfer between these projects, especially in terms of experiences gained by the single individuals. In order to exploit the full potential, single research activities in different projects should follow a clear strategy and build on each other's results. Exchanging experiences between different projects helps to prevent making the same mistakes in successive research activities and facilitate the identification of future research needs. This contribution aims to improve the knowledge transfer from the Semi40 project to future research activities, based on a workshop held with Semi40 industry experts. The goal is to incorporate experiences and suggestions for improvement from Semi40 in following projects to increase the efficiency of European research efforts. The workshop results are summarized and clustered in a graphical roadmap, enabling a temporal prioritization of identified topics.

### **Central Description of the Concept**

The results and research needs presented in this contribution are based on two central methods that build on each other. To start the session, the workshop participants performed a SWOT-analysis for the semiconductor industry in Europe. Applied on industry level, a SWOT-analysis is a strategic analytical tool for assessing strengths and weaknesses of the industry, analyzing opportunities available to the industry, as well as, threats faced by the industry. Widely used in businesses and research projects, a SWOT-analysis is a powerful tool that supports the user in assessing internal and external influences with a systematic approach, enabling the derivation of proper strategies. Even though a SWOT-analysis is not an essential requirement for the definition of roadmaps, the method was chosen with the intention to incite the creativity of the workshop participants. In addition, analyzing the semiconductor industry from today's perspective served as a starting point for the identification of future trends and challenges in the second part of the workshop. The second method used is roadmapping itself, which defines a strategic process of determining challenges, actions and possibly the resources needed to overcome these challenges. In order to support the participants in the development of the roadmaps, a template was provided, which is illustrated in Figure 1.

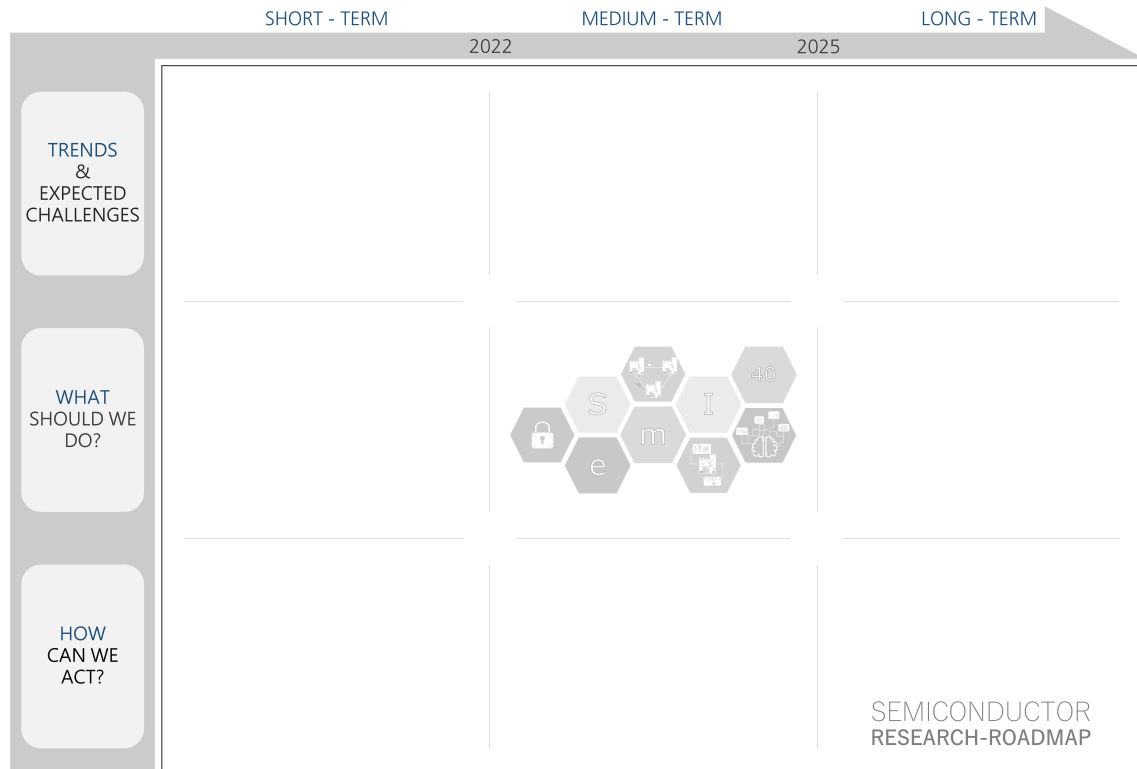


Figure 1: Roadmap workshop template

To create a roadmap, a time horizon has to be defined, which enables the chronological classification of the identified trends and challenges. The decision was taken to let the time horizon be defined by the industry experts themselves, since they could better estimate the relevance of trends in their industry. The fact that the experts defined a rather short time horizon is a clear indicator for the constant state of change in the semiconductor industry. While the short-term period, which is stated on the template, covers the first 2 to 3 years from now on, the medium-term period was defined to span the next 3 years, approximately until the year 2025. The template is further clustered into three sections, providing following key questions to the workshop participants:

- *Trends and expected challenges*
- *What should we do?*
- *How can we act?*

The three sections encouraged the industry experts to not only think about future trends and challenges, but to simultaneously consider possible solution approaches and how to translate these solutions into specific actions. This enabled a precise determination of research needs and further actions for the semiconductor industry.

#### *Preparation of Results*

Aiming to exploit the full creative potential of the workshop participants, no content restrictions were defined for the definition of future trends and challenges in the workshop. The participants were therefore free to choose whether to focus on specific issues that emerged during the Semi40 project or to address challenges on a global scale. The results are correspondingly diverse, spanning from general problems in research projects up to expected trends for the next years. To structure the wide range of results, the following sections are defined:

1. Global challenges
2. Challenges in the semiconductor industry
  - I. Need for collaboration – Business Ecosystems
  - II. Legal and ethical issues
  - III. Technology – New vs. Old
  - IV. Organization and management
3. Challenges in research projects
  - V. Data in research projects
  - VI. Transfer of research results

The first section summarizes current and future trends and challenges with implications on a global scale beyond the semiconductor industry. International efforts on both the political and academic level will be necessary to resolve these issues. The second section defines trends that will specifically affect the semiconductor industry. Based on these trends, relevant research topics and other challenges are defined that will need to be addressed in the in the upcoming years. Finally, the last section summarizes recommendations for future research projects that are derived from the specific insights and experiences made in the Semi40 project.

Each sub section is discussed in depth, presenting the challenge, an estimation of the temporal significance and possible solution approaches, basing on the presentations and discussions in the workshop. Identified research areas and other recommendations are highlighted at the end of each subsection. In order to provide an overview and a temporal prioritization, the most important trends, research needs and recommendations are summarized in a graphical roadmap in the final chapter.

## Results, Impact, Conclusion, Discussion

### 1. Global challenges

Some of the challenges addressed during the workshop are relevant beyond the semiconductor industry on a global scale. As these challenges will eventually have a major impact not only on the industry, but also on the society as a whole, finding solutions will have to take place primarily on a political and international level.

- **Demographic development:** The aging population is, according to the UN, poised to become one of the most significant social transformations and has great implications on all sectors of society. This trend is already leading to a decreased number of qualified experts in the labour market. In terms of knowledge management, there will be an increasing challenge in preparing the digital knowledge for future generations to use. Especially in times of an exploding amount of digital information, it is necessary to develop new methods to condense, preserve and to transfer the generated knowledge.
- **Limited Resources:** Diminishing resources and increasing environmental standards will force the manufacturing industry to search for more efficient production methods. Energy efficiency will be one of the main focuses, as the semiconductor industry is reliant on a high energy input. For example, electricity is required to operate the manufacturing process equipment, to maintain the cleanroom conditions, and to operate heating ventilation and air conditioning in the factory.
- **Market complexity:** Markets are becoming increasingly complex, not only because of more complex products, but also due to market dynamics, complex environments and

the intricacy of industry interaction. Identifying which kind of market will be feasible in such a complex environment will constitute a major challenge for the semiconductor industry. Continuous standardization will be a beneficial and necessary mechanism for the semiconductor industry to address this challenge.

- **Interdisciplinary work:** While globalisation is leading to an increased collaboration between people from different cultures, an increasing market complexity requires a stronger collaboration between experts from different disciplines. Interdisciplinary work is an important success factor in the 21st century, which is why there is a need to improve the framework conditions for interdisciplinary work. Regarding the educational system, many universities are still clustered into strict departments and separate curricula, preventing a knowledge transfer between different disciplines. A new educational system might be necessary in the near future to better prepare the next generations to the changing work requirements.

## 2. Challenges in the semiconductor industry

### I. Need for collaboration – Business Ecosystems

Increasing complexity in the semiconductor industry, caused by ongoing heterogeneous integration at the chip level, alongside other external influences, has a great influence on the role of collaboration between companies. In the medium-term, European semiconductor manufacturers will have to engage in horizontal and vertical collaboration activities with competitors, suppliers, customers and research centres in order to remain competitive. This need for collaboration could result in companies opening up their boundaries more willingly to engage in ad hoc partnerships, instead of waiting for mutual trust to be established over years. In terms of vertical integration along the value chain, the semiconductor industry currently faces the challenge of spatial separation. Many suppliers, as well as customers, are located in foreign countries, such as the US, China or South Korea, complicating potential collaboration opportunities. This challenge also affects the further development of production sites in Europe, as multinational customers with production facilities on other continents could demand shorter supply chains in the long-term. Future challenges for Europe are therefore twofold: First, it is necessary to intensify the research efforts in the field of faster and more efficient collaboration methods. Second, it is essential to increase the attractiveness of production locations in Europe by establishing whole Business Ecosystems. Having OEMs and Tier 1 suppliers in close proximity will be a key success factor to expand the semiconductor production in the European Union. Accordingly, a clear recommendation for future research projects is to focus the establishment of Business Ecosystems instead of funding single activities within one stage of the value chain. The suggestions for future research activities can be summarized as follows:

- Exploration of innovative and secure collaboration methods and tools to support a faster development of long-lasting cooperative relationships in the semiconductor industry.
- Integration of the whole semiconductor value chain and Business Ecosystems in EU funded research projects in order to strengthen the competitiveness of the European semiconductor industry from a global perspective.

### II. Legal and ethical issues

While Big Data and artificial intelligence (AI) are becoming more ubiquitous, more questions arise in terms of legal and ethical issues. For example, in the case of

autonomous vehicles, it has not yet been clarified who will be liable in the event of an accident. Is it fault of the company that manufactured the car or should vehicle owners be held liable? The growing dichotomy between AI and legal liability is currently seen as a black box that nobody feels responsible for. As chips are the primary enablers of autonomous driving, determining liability could eventually begin in the semiconductor industry. Through strict quality controls and following a zero-defect strategy, semiconductor manufacturers are already trying to protect themselves from all future eventualities. Yet more legal disputes are sure to arise, as long as further AI applications are developed every day. Since autonomous driving already is a huge market for the semiconductor industry, it is important to clarify these questions as soon as possible. With regard to future research projects, legal and ethical aspects should be addressed through greater interdisciplinary cooperation between the industry and legal experts. The following areas for action can be specified:

- Rapid clarification of legal issues to enable the implementation of autonomous driving pilots and other AI use cases in practice, thus ensuring a continuous technical progress.
- Integration of legal experts into future research projects in the field of AI. This will assure that solutions are not developed one-dimensionally from a technical perspective but with due regard to all external influencing factors.

### III. Technology – New vs. Old

#### *Technological trends*

Technological change is one of the most characteristic trends in the digital age. New technologies, user devices and methods of interaction between humans and machines appear nearly every day, promising changes to business and society. Digital trends, such as AI or Internet of Things (IoT) not only lead to increasing markets for the semiconductor industry, but also offer the opportunity for cost efficient production of semiconductors through self-organizing production systems. In order to exploit the advantages of these new technologies, it is important to pursue further interdisciplinary research projects, transferring theoretical concepts into practical solutions. The workshop brought up a number trends and technologies that will be relevant for the semiconductor industry in the coming years:

- Short-term: Gaining and using information from the ever-growing amount of data generated on factory level and the open world will remain at the core of industrial research activities. Structured data is the most important resource for simulations and ML (Machine Learning) algorithms, which then can produce models for predictive maintenance or fault detection systems. Together with Big Data Analytics and sensor development, the further validation and application of intelligent algorithms will be the key focuses in the short-term.
- Medium-term: A great number of data based decision-support systems and dashboards were developed over the past years and as a part of the Semi40 project. It is only a matter of time, until completely automated decision-making systems will be introduced in the industrial environment that will enable the human workforce to perform more knowledge intensive tasks. An increasing use of multimedia technology, such as VR (Virtual Reality) and AR (Augmented Reality), is also foreseeable in the near future. Several Use-Cases are already being researched in the fields of worker assistance, remote monitoring, collaboration and training that will soon reach market maturity.
- Long-term: Once the framework conditions are in place regarding sophisticated

algorithms, automated data provision and a secure IT infrastructure, the way will be paved for a full AI integration in manufacturing. A challenge which endangers the widespread use of AI, refers to the reliability and comprehensibility of information generated with AI. Humans tend to see AI as a “black box”, which produces an output that cannot be trusted, as the output is not traceable. The successful application of AI at the production line is therefore linked to the ability of making the output understandable to the user – a concept commonly referred to as “Explainable Artificial Intelligence (XAI)”.

#### *Old manufacturing equipment*

While digital technologies promise cost reduction, a faster time to market and an increasing customer awareness, manufacturers still have to cope with aging production plants that do not fit into the digital strategy. For cost reasons, it is not possible to simply replace old production plants, as the machines are designed to be used over many years. Manufacturers relying on machinery that was installed before concepts such as IoT and Big Data became popular consequently do not have the possibility to retrieve data from these machines. The development of extensions, data and controller interfaces is therefore a critical challenge that must be tackled in order to realize innovative and self-organizing production systems. From a product point of view, it may be necessary to switch to completely new manufacturing technologies in the future. The semiconductor industry has long followed Moore’s law, which predicts a doubling of transistors in a dense integrated circuit about every two years. Even though the end is not yet predictable, it is clear that Moore’s law will have a best before date. This end will force the semiconductor industry to search for new ways of producing chips with increasing performance. This will also lead to a change in the manufacturing equipment. The majority of European manufacturers still relies on old 100 nm technology to produce their chips. With a demand of increasing functionalities on the one hand, and structural limits in terms of layer thicknesses and channel lengths on the other hand, manufacturers will soon be required to switch to newer technologies. The same applies to the wafer thickness, which defines the number of chips that can be produced per wafer. While most manufacturers prefer 200 mm wafers, 300 mm wafers promise higher production yields, thus increasing the efficiency of chip production. It is very likely that increasing cost pressure will lead manufacturers to move to higher wafer diameters eventually. The following recommendations therefore arise in the context of old manufacturing equipment:

- Development of extensions, data and controller interfaces for old manufacturing equipment to allow for a total digitalization of semiconductor factories and thus enabling the implementation of innovative IoT concepts.
- Support European semiconductor manufacturers to make the technological switch to newer manufacturing technologies, due to changing wafer size and shrinking technology nodes. A successful transition will soon be a necessity to remain competitive in a global market.

#### IV. Organization and management

The integration of information and operational technology as well as the implementation of robotics and cyber-physical systems have provided industrial companies an increasing control of their production processes. In order to achieve the goal of a self-organizing and self-controlling factory, it is necessary to look beyond the shop floor and to start focusing organizational aspects. A majority of industrial companies still rely on



classical hierarchical production systems, which do not fit the strategy of Industry 4.0. A complete IoT and digitization strategy must therefore go hand in hand with the development and integration of a new process organization, including cooperate strategy, management system, work organization and human resource development. As seen from an optimization point of view, administrative processes offer the greatest potential increase in efficiency after production processes have been automated. The digitization of classic office processes through integration of automated decision-making systems and business process automation can be seen as the next big step to increase the competitiveness of industrial companies. This development will favor the emergence of new professions, where knowledge workers will have the chance to perform more challenging tasks. Future research projects in the semiconductor industry should therefore start addressing administrative and organizational aspects instead of focusing solely on technical issues.

- Development and application of new organizational methods, including corporate strategy, management system and work organization that are suitable for handling self-organizing teams and managing the increasing complexity in an Industry 4.0 environment.
- Development of new managerial approaches to secure full management support and commitment in times of digitization and automation.
- Focus shift from production process automation to business process automation. Identification of optimization potentials in administrative processes, which currently cause a large part of the company expenses.

### 3. Challenges in research projects

#### V. Data in research projects

The outcome of research projects in the manufacturing environment depends substantially on the quality of data exchange in the consortium. According to a popular concept in computer science – garbage in, garbage out (GIGO) – A high state of completeness, validity, consistency and accuracy of the underlying data is a key success factor in applying theoretical models and algorithms. This concept also applies to the outcomes of the Semi40 project, which had a strong focus on the development of data-driven applications such as predictive models, fleet management systems or fault detection algorithms. A differentiation between “data availability” and “data quality” is defined in this section to cluster the wide range of actions required.

##### *Data availability*

In the current situation, multiple factors hinder a proper data exchange between industrial partners and academic institutions. Among these factors, confidentiality regulations are particularly restrictive, forcing industrial partners to act with caution in order not to risk any loss of information. From a psychological point of view, single individuals might perceive the transfer of information as a loss of power and control, as the data might be used to demonstrate the inefficiency of current processes. From a technical perspective, an insufficient availability of data could also be caused by the use of old equipment and technologies in the production, which do not have proper data interfaces to generate the requisite information. The lack of data not only has a great impact on the outcomes of a project itself but also on the scope and the focus of future research activities. Intelligent algorithms, Big Data analytics and data-driven models developed in the course of a project cannot be validated properly, if data is not available in the

right quantity. This prevents future research activities to build on the achieved results and thus to make a step forward towards more interesting and innovative research areas. Accordingly, the workshop participants commonly agreed on the assumption, that the use of AI in production will not be achieved in the next years. In order to guarantee the progress of future research activities and the successful transfer of theory to practice, the issue of data availability needs to be tackled as soon as possible. Recommendations are:

- A clear definition of data requirements and a proper distinction between critical business data and data for the research project to be agreed on by all project partners.
- Research and development of methods for automated data collection in order to handle the constantly increasing amount of data.
- Enhanced communication, promotion and demonstration of Big Data benefits to industrial partners in order to incite the development of internal Big Data organization units.

#### *Data quality*

The quality of issued data is equally important as the right quantity. Data which has a lack of completeness, consistency, accuracy, validity or timeliness, can lead to faulty model outputs and subsequently to wrong decisions. From all the challenges stated during the workshop, insufficient data quality was claimed to be the main reason that Use-Cases in the Semi40 project could not be fully exploited. Instead of working on Use-Case specific solutions, academic institutions had to spend large part of the time available to discuss and to clean up data, in order to achieve the minimum requirements. The growing challenge of ensuring a sufficient data quality not only applies to the Semi40 project but to all research projects with a high data and information exchange. Dealing with this concern is therefore a crucial step to increase the efficiency and to improve the results of future research activities. A number of proposals and possible solutions were presented during the workshop:

- Development of data exchange standards with regard to the completeness, consistency, accuracy, validity and timeliness of the issued data. This should help to standardize the data exchange process in order to reduce the amount of time required for data cleansing activities.
- Further development of technical solutions for secure and more efficient data exchange (e.g. based on semantic web, Blockchain technology etc.).
- Establishment of a fixed position in large-scale research projects in charge of the complete data exchange and data cleansing process. This “Data Cleansing House”, performed by a research institution or a university, could positively influence the efficiency by allowing other academic partners to focus on research activities.

#### VI. Transfer of research results

The success of research projects is often measured by the extent to which research findings are translated into practical applications. Due to various reasons, a successful transfer does not always take place. In particular regarding cases, where research findings need to be integrated into running production, many projects are doomed to failure, as managers do not want to introduce losses by interrupting the production. In these special cases, it is necessary to align the strategy of the research project with the corporate strategy in advance, in order to secure a top management support in the implementation phase. Another reason for a difficult translation of research findings

relates to the lack of standards. For example, in IT security, many different security protocols are currently available, which provide a secure delivery of data between two parties. In the Semi40 project, a few solutions in the field of IT security were elaborated, which have to be aligned to a certain protocol in order to be implemented. Even though the technical IT solutions are already developed, many companies choose to wait until a certain standard is evolved among the various security protocols, before integrating the solutions. This causes a delay of many month, even years, before research findings are translated into practice. The following recommendations should therefore be considered, in order to make a better use of research results in future research projects:

- Better alignment of project content to the corporate strategies of the partners prior to the start of the project. This will enhance the significance of the research projects and thereby help to secure the necessary top management support in the implementation phase of project results.
- Encouragement of project partners to act proactively in the case of unclear standards. The utilization of project results should not be dependent on the emergence of one clear standard. If a certain option is mature enough, project partners should have the courage to build upon this option in order to set a standard themselves.

### Discussion

The graphical roadmap summarizes the most important trends and recommendations for research activities after Semi40 that were identified and discussed during the workshop. The results are arranged according to their expected significance over the coming years. The timeline does not define the starting point for research activities. It rather expresses a realistic expectation when stated events and research areas will become relevant for the semiconductor industry. For example, in the case of AI, a full integration in manufacturing is not realistic before other current challenges are dealt with. The roadmap classifies each topic into the time intervals: Short-term, medium-term and long-term. A specification of dates is not appropriate, since the arrangement is based on a rough estimate.

Figure 2 clearly illustrates, that data will continue to be a critical topic for the coming years. Basic challenges related to data availability and data quality need to be addressed as soon as possible, before more complex applications such as augmented reality or AI will have a major influence in manufacturing. Tackling the challenge of insufficient data quality through standardization will also be critical to enhance the efficiency in data-driven research activities, while simultaneously improving the exploitation of results. Apart from technical developments, it is necessary to cope with questions and issues that are raised by these developments. This applies, in particular, to legal and ethical issues, which continuously emerge with the development of new AI applications. Unresolved legal issues hinder the implementation of pilot projects and thus prevent technical progress. Funding interdisciplinary research projects that integrate technical specialists as well as legal experts will help to solve such issues in future research activities. Achieving full digitization will finally require a shift towards more holistic research projects, moving away from a pure technical perspective. This will require both an organizational redesign as well as the integration of the whole semiconductor value chain in research projects. Supporting collaboration and the development of Business Ecosystems will be a major success factor to keep the production in Europe from a long-term perspective.

The Semi40-project has produced many innovative and recognised results in the areas of smart production and cyber-physical production systems. Apart from these physical outcomes, the experiences and insights gained over the period of the project are essential results that can benefit future research projects. This contribution summarizes the experiences made by industry experts during

Semi40, highlighting suggestions for improvement and future research needs. The contribution can therefore serve as a starting point for further research activities, aiming to improve future research projects in the semiconductor industry, both in terms of content and methodology.

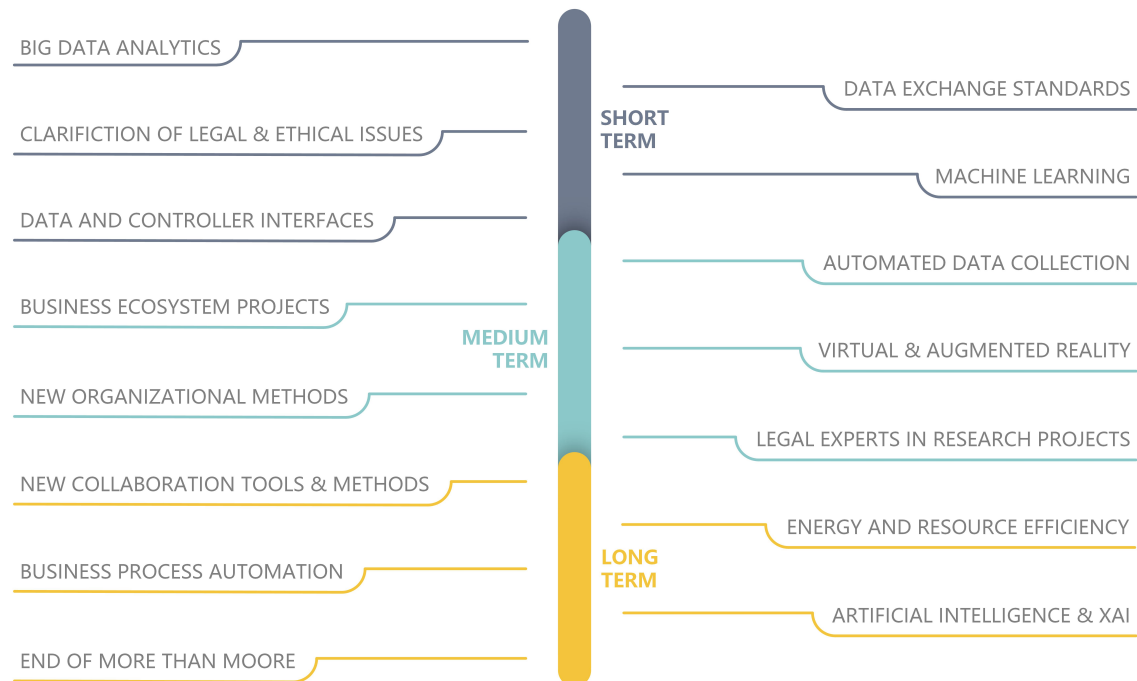


Figure 2: Semiconductor research roadmap after SEMI40

#### References *References<sub>semi40,roadmap</sub>*

##### About Authors

**Dipl. -Ing. Alessandro Sala** holds a Master's degree in "Mechanical Engineering – Management" from the Technical University of Vienna, with his thesis evaluating the impacts of Digitalization on a company's innovation performance. He joined Fraunhofer Austria Research GmbH in 2017 and is currently working as a research associate in the division of Production and Logistics Management, with special interests in Technology and Innovation Management and Data Science.

**Marius Lütkemeyer M.Sc.** finished his Master studies in "International Industrial Engineering" at the University of Applied Science Technikum Wien. The main focus of these studies was production management. He has been working at Fraunhofer Austria Research GmbH as a research associate since 2017. His areas of expertise are factory planning and factory organization. In the Semi40 project, Marius Lütkemeyer was responsible for the evaluation of the impact by the industry 4.0 use cases.

**Andreas Felsberger, M.Sc.** finished his Master studies in "General Management" at the University of Klagenfurt. He joined University of Klagenfurt in 2016 and is currently working as a research associate in the department of production management and logistics. Andreas is currently working on his doctoral thesis, which was largely made possible by Semi40. His research focus lies in the field of impact evaluation of digitization strategies and new technologies.





## Bibliography

- [1] Federica Acerbi et al. “Accuracy and Robustness Against Covariate Shift of Water Chiller Models”. In: *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*. IEEE. 2018, pages 809–816 (cited on page 82).
- [2] P Adolphs et al. “Referenzarchitekturmodell Industrie 4.0 (RAMI4. 0)”. In: *VDI/VDE Statusreport* (2015) (cited on page 30).
- [3] John Aldrich et al. “Correlations genuine and spurious in Pearson and Yule”. In: *Statistical science* 10.4 (1995), pages 364–376.
- [4] B V Almende, T Benoit, and R Titouan. *Package ‘visNetwork’: Network Visualization using ‘vis.js’ Library*. <https://cran.r-project.org/web/packages/visNetwork/index.html>. Accessed on 29.01.2019 (cited on page 49).
- [5] Khaled Alsabti, Sanjay Ranka, and Vineet Singh. “An efficient k-means clustering algorithm”. In: (1997).
- [6] Sarabjot S Anand and Alex G Büchner. *Decision support using data mining*. Financial Times Management, 1998 (cited on page 65).
- [7] *Apache Shiro*. <https://shiro.apache.org>. Accessed: 16-01-2019. 2019 (cited on page 116).
- [8] David Arthur and Sergei Vassilvitskii. “k-means++: The advantages of careful seeding”. In: *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics. 2007, pages 1027–1035.
- [9] Sugato Bagchi et al. “A full-factory simulator as a daily decision-support tool for 300mm wafer fabrication productivity”. In: *2008 Winter Simulation Conference*. IEEE. 2008, pages 2021–2029 (cited on page 107).
- [10] John J Bartko. “The intraclass correlation coefficient as a measure of reliability”. In: *Psychological reports* 19.1 (1966), pages 3–11.

- [11] David Basin, Felix Klaedtke, and Samuel Müller. “Monitoring security policies with metric first-order temporal logic”. In: *Proceedings of the 15th ACM symposium on Access control models and technologies*. ACM. 2010, pages 23–34 (cited on page 8).
- [12] SS Baskar, L Arockiam, and S Charles. “A systematic approach on data pre-processing in data mining”. In: *Compusoft* 2.11 (2013), page 335 (cited on page 67).
- [13] Dennis Bauer, Andreas Schlereth, and Florian Maier. *Simulation-Assisted Decision Making for Supply Chain Disruptions in Production Control*. 2018 (cited on page 39).
- [14] Dennis Bauer et al. “Concept and Possible Application of an Automated Framework to Influence Production Dispatch Based on Supply Chain Events”. In: ().
- [15] Thomas Bauernhansl, Anja Schatz, and Jens Jäger. “Komplexität bewirtschaften–Industrie 4.0 und die Folgen”. In: *ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb* 109.5 (2014), pages 347–350 (cited on page 36).
- [16] Asa Ben-Hur et al. “Support vector clustering”. In: *Journal of machine learning research* 2.Dec (2001), pages 125–137.
- [17] Dinkar N Bhat and Shree K Nayar. “Ordinal measures for image correspondence”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.4 (1998), pages 415–423 (cited on page 99).
- [18] *BI and ETL Process Management Pain Points*. [https://www.cisco.com/c/dam/en/us/products/collateral/cloud-systems-management/tidal-enterprise-scheduler/whitepaper\\_c11-633329.pdf](https://www.cisco.com/c/dam/en/us/products/collateral/cloud-systems-management/tidal-enterprise-scheduler/whitepaper_c11-633329.pdf). Retrieved: 28-05-2013. 2010 (cited on pages 127, 130).
- [19] Ani Bicaku et al. “Harmonized monitoring for high assurance clouds”. In: *2016 IEEE International Conference on Cloud Engineering Workshop (IC2EW)*. IEEE. 2016, pages 118–123 (cited on page 8).
- [20] Ani Bicaku et al. “Towards trustworthy end-to-end communication in industry 4.0”. In: *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*. IEEE. 2017, pages 889–896 (cited on page 7).
- [21] Ani Bicaku et al. “Monitoring Industry 4.0 applications for security and safety standard compliance”. In: *2018 IEEE Industrial Cyber-Physical Systems (ICPS)*. IEEE. 2018, pages 749–754 (cited on page 8).
- [22] Jürgen Bischoff. *Studie" Erschließen der Potenziale der Anwendung von'Industrie 4.0'im Mittelstand"*. agiplan GmbH, 2015 (cited on page 36).
- [23] Andrew Blake, Pushmeet Kohli, and Carsten Rother. *Markov random fields for vision and image processing*. Mit Press, 2011.
- [24] Andrei Borshchev. “Multi-method modeling”. In: *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*. IEEE Press. 2013, pages 4089–4100.
- [25] Alfons Botthof and Ernst Andreas Hartmann. *Zukunft der Arbeit in Industrie 4.0*. Springer Vieweg Berlin, 2015.
- [26] Peter Brandl et al. “Assist 4.0-Datenbrillen-Assistenzsysteme im Praxiseinsatz.” In: *Mensch & Computer Workshopband 2014* (2014), pages 259–264.
- [27] L Breiman et al. “Classification and regression trees (Wadsworth, Belmont, CA)”. In: *ISBN-13* (1984), pages 978–0412048418.



- [28] L Breiman et al. *Package 'randomForest': Breiman and Cutler's Random Forests for Classification and Regression*. <https://cran.r-project.org/web/packages/randomForest/>. Accessed on 29.01.2019.
- [29] Roland Buchner et al. "Designing for the Factory: UX Prototyping for the Cleanroom". In: *INTERACT 2015 Adjunct Proceedings: 15th IFIP TC. 13 International Conference on Human-Computer Interaction 14-18 September 2015, Bamberg, Germany*. Volume 22. 2015, page 145.
- [30] Peter Cabena et al. *Discovering data mining: from concept to implementation*. Prentice Hall PTR New Jersey, 1997 (cited on page 65).
- [31] Gianni Campatelli, Alexander Richter, and Alexander Stocker. "Participative knowledge management to empower manufacturing workers". In: *International Journal of Knowledge Management (IJKM)* 12.4 (2016), pages 37–50.
- [32] David Cemernek, Heimo Gursch, and Roman Kern. "Big data as a promoter of industry 4.0: Lessons of the semiconductor industry". In: *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*. IEEE. 2017, pages 239–244.
- [33] Steve Chadwick et al. "it@ intel". In: (2016) (cited on page 127).
- [34] Pete Chapman et al. "The CRISP-DM process model". In: *The CRIP-DM Consortium* 310 (1999) (cited on page 65).
- [35] Toly Chen. "Forecasting the yield of a semiconductor product with a collaborative intelligence approach". In: *Applied Soft Computing* 13.3 (2013), pages 1552–1560 (cited on page 48).
- [36] Toly Chen. "An ANN approach for modeling the multisource yield learning process with semiconductor manufacturing as an example". In: *Computers & Industrial Engineering* 103 (2017), pages 98–104 (cited on page 48).
- [37] Raman Chitkara et al. "The Internet of Things: The next growth engine for the semiconductor industry". In: *PwC*. Volume 1. 2015, page 36.
- [38] Charles K Chui and HN Mhaskar. "MRA contextual-recovery extension of smooth functions on manifolds". In: *Applied and Computational Harmonic Analysis* 28.1 (2010), pages 104–113.
- [39] K Cios and L Kurgan. *Trends in Data Mining and Knowledge Discovery. Advanced Techniques in Knowledge Discovery and Data Mining*. N. Pal and L. Jain. 2005 (cited on page 65).
- [40] Krzysztof J Cios et al. "The knowledge discovery process". In: *Data Mining*. Springer. 2007, pages 9–24 (cited on page 65).
- [41] Michael Clark. "A comparison of correlation measures". In: *Center for Social Research, University of Notre Dame* 4 (2013).
- [42] Corinna Cortes and Vladimir Vapnik. "Support-vector networks". In: *Machine learning* 20.3 (1995), pages 273–297.
- [43] Louise Corti et al. *Managing and sharing research data: a guide to good practice*. Sage, 2014 (cited on page 59).
- [44] Scott Deerwester et al. "Indexing by latent semantic analysis". In: *Journal of the American society for information science* 41.6 (1990), pages 391–407 (cited on page 68).
- [45] Andrea Denger et al. "Organisationaler Wandel durch die Emergenz Cyber-Physikalischer Systeme: Die Fallstudie AVL List GmbH". In: *HMD Praxis der Wirtschaftsinformatik* 51.6 (2014), pages 827–837.

- [46] Friedrich Dornbusch. "Global competition in microelectronics industry from a European perspective: Technology, markets and implications for industrial policy". In: ().
- [47] Allen B Downey. *Think stats*. " O'Reilly Media, Inc.", 2011.
- [48] *EclipseLink*. <http://www.eclipse.org/eclipselink>. Accessed: 16-01-2019. 2019 (cited on page 116).
- [49] *Energy Efficient Computing*. <https://sing.stanford.edu/site/projects/9> (cited on page 127).
- [50] Martin Ester et al. "A density-based algorithm for discovering clusters in large spatial databases with noise." In: *Kdd*. Volume 96. 34. 1996, pages 226–231.
- [51] Manolis Falelakis et al. "Connecting Living Rooms: An Experiment In Orchestrated Video Communication". In: *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*. TVX'16. Chicago, Illinois, USA: ACM, 2016, pages 49–58. ISBN: 978-1-4503-4067-0. DOI: 10.1145/2932206.2932215. URL: <http://doi.acm.org/10.1145/2932206.2932215> (cited on page 145).
- [52] Christos Faloutsos and Vasileios Megalooikonomou. "On data mining, compression, and kolmogorov complexity". In: *Data mining and knowledge discovery* 15.1 (2007), pages 3–20 (cited on page 126).
- [53] Aziza Faruqi et al. "Test wafer management and automated wafer sorting". In: *2008 IEEE/SEMI Advanced Semiconductor Manufacturing Conference*. IEEE. 2008, pages 322–326 (cited on page 91).
- [54] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases". In: *AI magazine* 17.3 (1996), page 37 (cited on page 65).
- [55] Patrick Ferland and Andre Labonte. "Manufacturing efficiency improvement through automation of test wafer procedures". In: *2004 IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop (IEEE Cat. No. 04CH37530)*. IEEE. 2004, pages 321–324.
- [56] Stephen Few. "Information dashboard design". In: (2006) (cited on page 75).
- [57] Luigi Fortuna et al. *Soft sensors for monitoring and control of industrial processes*. Springer Science & Business Media, 2007.
- [58] Bryce Foster et al. "Simulation of test wafer consumption in a semiconductor facility". In: *IEEE/SEMI 1998 IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop (Cat. No. 98CH36168)*. IEEE. 1998, pages 298–302 (cited on page 91).
- [59] John W Fowler, Lars Mönch, and Thomas Ponsignon. "DISCRETE-EVENT SIMULATION FOR SEMICONDUCTOR WAFER FABRICATION FACILITIES: A TUTORIAL." In: *International Journal of Industrial Engineering* 22.5 (2015).
- [60] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Volume 1. 10. Springer series in statistics New York, 2001.
- [61] S Fritsch et al. *Package 'neuralnet: Network Visualization using 'vis.js' Library*. <https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf>. Accessed on 30.01.2019 (cited on page 51).
- [62] G Gautreau. *jQRangeSlider*. <http://ghusse.github.io/jQRangeSlider>. Accessed: 16-01-2019. 2019 (cited on page 116).

- [63] William W. Gaver. "Technology Affordances". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI'91. New Orleans, Louisiana, USA: ACM, 1991, pages 79–84. ISBN: 0-89791-383-3. DOI: 10.1145/108844.108856. URL: <http://doi.acm.org/10.1145/108844.108856> (cited on page 148).
- [64] John Gerring. "What is a case study and what is it good for?" In: *American political science review* 98.2 (2004), pages 341–354.
- [65] *Global Semiconductor Market Forecast 2017-2024*. <https://www.reportlinker.com/p04595244/Global-Semiconductors-Market-Forecast.html>. Retrieved: December 10, 2018, from [www.reportlinker.com](http://www.reportlinker.com). Oct. 2017 (cited on page 76).
- [66] Andreas Gödl and Peter Brandl. "Multimodale Interaktion mit HMIs in der Smart Factory". In: *Mensch und Computer 2017 - Workshopband*. Edited by Manuel Burghardt et al. Regensburg: Gesellschaft für Informatik e.V., 2017.
- [67] Leo A Goodman and William H Kruskal. "Measures of association for cross classifications". In: *Measures of association for cross classifications*. Springer, 1979, pages 2–34.
- [68] Geoffrey Gordon. "A general purpose systems simulation program". In: *Proceedings of the December 12-14, 1961, eastern joint computer conference: computers-key to total systems control*. ACM. 1961, pages 87–104 (cited on page 106).
- [69] JM Gordon and Kim Choon Ng. "Thermodynamic modeling of reciprocating chillers". In: *Journal of Applied Physics* 75.6 (1994), pages 2769–2774 (cited on page 81).
- [70] Benjamin Götz et al. "Challenges of production microservices". In: *Procedia CIRP* 67.1 (2018), pages 167–172 (cited on page 39).
- [71] Larry Greenemeier. *How Close Are We—Really—to Building a Quantum Computer?* <https://www.scientificamerican.com/article/how-close-are-we-really-to-building-a-quantum-computer>. Retrieved from Scientific American. 2018 (cited on page 129).
- [72] L.A. Grieco et al. *IoT-Aided Robotics Applications: Technological Implications, Target Domains and Open Issues*. Elsevier, 2014, pages 32–47 (cited on page 17).
- [73] Michael Grieves. "Digital twin: Manufacturing excellence through virtual factory replication". In: *White paper* (2014), pages 1–7 (cited on page 108).
- [74] Frank E Grubbs. "Procedures for detecting outlying observations in samples". In: *Technometrics* 11.1 (1969), pages 1–21.
- [75] Kurt Gruber, Christian Pophal, and Hans Ehm. "Infineon: Integrated Supply Chain Architecture to Support Sustainability". In: *Sustainable Value Chain Management*. Springer, 2015, pages 381–391 (cited on page 36).
- [76] *Guideline Supply Chain Management in Electronics Manufacturing*. <https://www.zvei.org/en/press-media/publications/guideline-supply-chain-management-in-electronics-manufacturing>. 2014 (cited on page 39).
- [77] Isabelle Guyon and André Elisseeff. "An introduction to variable and feature selection". In: *Journal of machine learning research* 3.Mar (2003), pages 1157–1182.
- [78] S. Haas, T. Ulz, and C. Steger. "Secured Offline Authentication on Industrial Mobile Robots using Biometric Data". In: *RoboCup International Symposium 2017*. Springer. 2017, pages 143–155 (cited on pages 18, 21).
- [79] S. Haas, T. Ulz, and C. Steger. "Secured Action Authorization for Industrial Mobile Robots". In: *IEEE Industrial Cyber-Physical Systems (ICPS)*. IEEE. 2018, pages 806–811 (cited on pages 18, 20, 21).

- [80] J. Haid. *Hardware-based solutions secure machine identities in smart factories*. Boards Solutions, 2016, pages 10–13 (cited on pages 17, 21).
- [81] Christian Hammel et al. “Empowering existing automated material handling systems to rising requirements: FA: Factory automation”. In: *2016 27th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*. IEEE. 2016, pages 87–93 (cited on page 91).
- [82] Arun Hampapur and Ruud M Bolle. “Comparison of distance measures for video copy detection”. In: *null*. IEEE. 2001, page 188 (cited on page 94).
- [83] Lea Hannola et al. “Empowering production workers with digitally facilitated knowledge processes—a conceptual framework”. In: *International Journal of Production Research* 56.14 (2018), pages 4729–4743.
- [84] Robert L Harrison. “Introduction to monte carlo simulation”. In: *AIP conference proceedings*. Volume 1204. 1. AIP. 2010, pages 17–21 (cited on page 106).
- [85] Lance Hayden. *IT security metrics: A practical framework for measuring security & protecting data*. Volume 396. McGraw Hill New York, 2010 (cited on page 8).
- [86] M. Hermann, T. Pentek, and B Otto. “Design Principles for Industrie 4.0 Scenarios”. In: *49th Hawaii International Conference on System Sciences (HICSS)*. IEEE. 2016, pages 3928–3937 (cited on page 17).
- [87] Mario Hermann, Tobias Pentek, and Boris Otto. “Design principles for industrie 4.0 scenarios”. In: *2016 49th Hawaii international conference on system sciences (HICSS)*. IEEE. 2016, pages 3928–3937.
- [88] Sandra Hirsh, Abigail Sellen, and Nancy Brokopp. “Why HP people do and don’t use videoconferencing systems”. In: *Hewlett-Packard Laboratories* (2005) (cited on page 147).
- [89] HL Horak. “DOE-2 reference manual”. In: (1979) (cited on page 82).
- [90] Jer-Wei Hsu et al. “Test wafer control system in 300 mm fab”. In: *2004 Semiconductor Manufacturing Technology Workshop Proceedings (IEEE Cat. No. 04EX846)*. IEEE. 2004, pages 33–36 (cited on page 91).
- [91] Xian-Sheng Hua, Xian Chen, and Hong-Jiang Zhang. “Robust video signature based on ordinal measure”. In: *2004 International Conference on Image Processing, 2004. ICIP’04*. Volume 1. IEEE. 2004, pages 685–688 (cited on pages 98, 99).
- [92] Yilin Huang. “Automated simulation model generation”. In: (2013) (cited on page 108).
- [93] *International Technology Roadmap for Semiconductors*. [https://www.dropbox.com/sh/vxigcu48nfe4t81/AACuMvZEhlpeQ6G8miYFCSEJa?dl=0&preview=1\\_Executive+Summary.pdf](https://www.dropbox.com/sh/vxigcu48nfe4t81/AACuMvZEhlpeQ6G8miYFCSEJa?dl=0&preview=1_Executive+Summary.pdf). 2001 (cited on page 36).
- [94] Ivar Jacobson. *Object-oriented software engineering: a use case driven approach*. Pearson Education India, 1993.
- [95] Gareth James et al. *An introduction to statistical learning*. Volume 112. Springer, 2013 (cited on pages 48, 50).
- [96] *Java Card Technology*. <https://www.oracle.com/technetwork/java/embedded/javacard/overview/index.html>. 2018 (cited on page 22).
- [97] I T Jolliffe. *Principal Component Analysis*. Volume 2. Springer, 2002 (cited on pages 68, 69).

- [98] Erland Jonsson. “Towards an integrated conceptual model of security and dependability”. In: *First International Conference on Availability, Reliability and Security (ARES’06)*. IEEE. 2006, 8–pp (cited on page 8).
- [99] Petr Kadlec, Bogdan Gabrys, and Sibylle Strandt. “Data-driven soft sensors in the process industry”. In: *Computers & chemical engineering* 33.4 (2009), pages 795–814.
- [100] Henning Kagermann, Wolfgang Wahlster, and Johannes Helbig. “Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0–Abschlussbericht des Arbeitskreises Industrie 4.0”. In: *Frankfurt am Main* (2013).
- [101] Rene Kaiser. “Opportunities and Challenges of Video Content and Video Technology in Smart Factories”. In: *Beiträge zum Usability Day XVI – Assistenztechnologien in der Arbeitswelt*. uDay XVI. Dornbirn, Austria: Pabst Science Publishers, 2018, pages 91–101. URL: <https://www.fhv.at/forschung/nutzerzentrierte-technologien/uday/>.
- [102] Rene Kaiser. “Towards Applying the Virtual Director Concept to 360 Degree Video Content”. In: *Adjunct Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video (TVX’18), Workshop Day on Virtual Reality, published via figshare*. 2018. DOI: 10.6084/m9.figshare.6526484.v11. URL: [https://figshare.com/articles/WS\\_Towards\\_Applying\\_the\\_Virtual\\_Director\\_Concept\\_to\\_360\\_Degree\\_Video\\_Content/6526484/1](https://figshare.com/articles/WS_Towards_Applying_the_Virtual_Director_Concept_to_360_Degree_Video_Content/6526484/1).
- [103] Rene Kaiser and Ferdinand Fuhrmann. “Multimodal Interaction for Future Control Centers: Interaction Concept and Implementation”. In: *Proceedings of the 2014 Workshop on Roadmapping the Future of Multimodal Interaction Research Including Business Opportunities and Challenges*. RFMIR ’14. Istanbul, Turkey: ACM, 2014, pages 47–51. ISBN: 978-1-4503-0615-7. DOI: 10.1145/2666253.2666255. URL: <http://doi.acm.org/10.1145/2666253.2666255> (cited on page 146).
- [104] Rene Kaiser and Wolfgang Weiss. “Media Production, Delivery and Interaction for Platform Independent Systems: Format-Agnostic Media”. In: edited by Oliver Schreer et al. Wiley, 2014. Chapter Virtual Director, pages 209–259. ISBN: 978-1-118-60533-2 (cited on page 145).
- [105] Rene Kaiser et al. “Collaborating in a Research and Development Project: Knowledge Protection Practices applied in a Co-opetitive Setting”. In: *10th Conference Professional Knowledge Management, Data-Driven Knowledge Management workshop*. proWM’19. Potsdam, DE, 2019.
- [106] Michael Steinbach George Karypis, Vipin Kumar, and Michael Steinbach. “A comparison of document clustering techniques”. In: *TextMining Workshop at KDD2000 (May 2000)*. 2000.
- [107] John D Kelleher, Brian Mac Namee, and Aoife D’arcy. *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT Press, 2015 (cited on page 25).
- [108] Karl G Kempf. “Simulating semiconductor manufacturing systems: successes, failures, and deep questions”. In: *Proceedings of the 28th conference on Winter simulation*. IEEE Computer Society. 1996, pages 3–11.
- [109] Maurice G Kendall and J Dickinson Gibbons. “Rank correlation methods. Ed”. In: *Edward Arnold* (1990).

- [110] Dongil Kim et al. “Machine learning-based novelty detection for faulty wafer detection in semiconductor manufacturing”. In: *Expert Systems with Applications* 39.4 (2012), pages 4075–4083 (cited on page 48).
- [111] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *arXiv preprint arXiv:1312.6114* (2013).
- [112] Mario Kleindienst et al. “Industry 4.0: What workers need and what ICT can give-an analysis”. In: *i-Know 2016: 16th International Conference on Knowledge Technologies and Data-driven Business*. 2016.
- [113] Ned Kock. “The Psychobiological Model: Towards a New Theory of Computer-Mediated Communication Based on Darwinian Evolution”. In: *Organization Science* 15.3 (June 2004), pages 327–348. ISSN: 1526-5455. DOI: 10.1287/orsc.1040.0071. URL: <http://dx.doi.org/10.1287/orsc.1040.0071>.
- [114] Miloš Kravčák et al. “Adaptive Video Techniques for Informal Learning Support in Workplace Environments”. In: *Emerging Technologies for Education*. Edited by Ting-Ting Wu et al. Cham: Springer International Publishing, 2017, pages 533–543. ISBN: 978-3-319-52836-6.
- [115] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems*. 2012, pages 1097–1105.
- [116] Max Kuhn and Kjell Johnson. *Applied predictive modeling*. Volume 26. Springer, 2013 (cited on pages 25, 48, 50).
- [117] Sherry Lake. *Best Practices Data Management*. <http://www.slideshare.net/shlake/best-practices-data-management>. 2015.
- [118] Averill M Law and Michael G McComas. “Simulation-based optimization”. In: *Proceedings of the Winter Simulation Conference*. Volume 1. IEEE. 2002, pages 41–44 (cited on page 106).
- [119] Charles L Lawson and Richard J Hanson. *Solving least squares problems*. Volume 15. Siam, 1995 (cited on pages 53, 56).
- [120] L Lazlo. *QR Code scanner*. <https://webqr.com>. Accessed: 16-01-2019. 2019 (cited on page 116).
- [121] Yann LeCun et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pages 2278–2324.
- [122] Jay Lee, Hung-An Kao, and Shanhu Yang. “Service innovation and smart analytics for industry 4.0 and big data environment”. In: *Procedia Cirp* 16 (2014), pages 3–8.
- [123] Jeff Leek, Leonardo Collado-Torres, and Nick Reich. *How to share data with a statistician*. <https://github.com/jtleek/datasharing>. 2016.
- [124] C. Lesjak, D. Hein, and J. Winter. “Hardware-security technologies for industrial IoT: TrustZone and security controller”. In: *41st Annual Conference of the IEEE Industrial Electronics Society*. IEEE. 2015, pages 2589–2595 (cited on page 21).
- [125] Ker-Chau Li. “Sliced inverse regression for dimension reduction”. In: *Journal of the American Statistical Association* 86.414 (1991), pages 316–327.
- [126] Bao Lin et al. “A systematic approach for soft sensor development”. In: *Computer Aided Chemical Engineering*. Volume 20. Elsevier, 2005, pages 1147–1152.

- [127] James T Lin and Chien-Ming Chen. “Simulation optimization approach for hybrid flow shop scheduling problem in semiconductor back-end manufacturing”. In: *Simulation Modelling Practice and Theory* 51 (2015), pages 100–114 (cited on page 48).
- [128] Lukas Lingitz et al. “Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer”. In: *PROCEDIA CIRP* 72 (2018), pages 1051–1056 (cited on pages 26, 48).
- [129] Chuen-Shiuan Liou et al. “Capacity forecast model for control and dummy wafers”. In: *ISSM 2005, IEEE International Symposium on Semiconductor Manufacturing, 2005*. IEEE. 2005, pages 123–125 (cited on page 91).
- [130] H. Lödding. *Verfahren der Fertigungssteuerung: Grundlagen, Beschreibung, Konfiguration (3rd edition)*. Springer Vieweg, 2016 (cited on page 36).
- [131] Hermann Lödding. *Verfahren der Fertigungssteuerung*. Volume 2. Springer, 2005.
- [132] G Luhn. “Closing the gap: Real-time production information system as based on a holistic information model HIM”. SYSTEMA concept paper. 2017 (cited on pages 126–128).
- [133] Laurens van der Maaten and Geoffrey Hinton. “Visualizing data using t-SNE”. In: *Journal of machine learning research* 9.Nov (2008), pages 2579–2605 (cited on page 68).
- [134] Alexander Maier, Tim Tack, and Oliver Niggemann. “Visual Anomaly Detection in Production Plants.” In: *ICINCO (1)*. 2012, pages 67–75 (cited on page 98).
- [135] Silia Maksuti et al. “Towards flexible and secure end-to-end communication in industry 4.0”. In: *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*. IEEE. 2017, pages 883–888 (cited on page 9).
- [136] Jonathan I Maletic and Andrian Marcus. “Data cleansing: A prelude to knowledge discovery”. In: *Data Mining and Knowledge Discovery Handbook*. Springer, 2009, pages 19–32 (cited on page 67).
- [137] MariaDB. <https://mariadb.org>. Accessed: 16-01-2019. 2019 (cited on page 116).
- [138] Lothar März et al. *Simulation und Optimierung in Produktion und Logistik: Praxisorientierter Leitfaden mit Fallbeispielen*. Springer-Verlag, 2010.
- [139] Stephen C Mathewson. “The application of program generator software and its extensions to discrete event simulation modeling”. In: *IIE transactions* 16.1 (1984), pages 3–18 (cited on page 108).
- [140] Matlab MathWorks. *MATLAB für Künstliche Intelligenz*. <https://de.mathworks.com/>. Accessed on 29.01.2019 (cited on page 48).
- [141] Y Meyer. “Wavelets and Operators, Analysis at Urbana, vol. 1, edited by E. Berkson, NT Peck and J. Uhl”. In: *London Math. Society, Lecture Notes Series* 137 (1989) (cited on page 69).
- [142] Rakesh Mohan. “Video sequence matching”. In: *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP’98 (Cat. No. 98CH36181)*. Volume 6. IEEE. 1998, pages 3697–3700 (cited on page 99).
- [143] L. Mönch, J. W. Fowler, and S. J. Mason. *Production planning and control for semiconductor wafer fabrication facilities: Modeling, analysis, and systems. Operations Research / Computer Science Interfaces Series*. Volume 52. 2013 (cited on page 38).
- [144] Lars Mönch, John W Fowler, and Scott J Mason. *Production planning and control for semiconductor wafer fabrication facilities: modeling, analysis, and systems*. Volume 52. Springer Science & Business Media, 2012 (cited on page 36).



- [145] Alessio Mosca et al. “Optimal Wafer Dispatching based on Dynamic Programming”. In: *18th European Advanced Process Control and Manufacturing Conference*. 2018 (cited on page 86).
- [146] James Moyne. “Run-to-run control in semiconductor manufacturing”. In: *Encyclopedia of Systems and Control* (2015), pages 1248–1254 (cited on page 38).
- [147] Sonja Grgic—Mislav Grgic—Marta Mrak et al. “Reliability of objective picture quality measures”. In: *Journal of Electrical Engineering* 55.1-2 (2004), pages 3–10 (cited on page 94).
- [148] Tamara Munzner. *Visualization analysis and design*. AK Peters/CRC Press, 2014 (cited on page 72).
- [149] S. Myagmar, A.J. Lee, and W. Yurcik. *Threat Modeling as a Basis for Security Requirements*. Citeseer, 2005, pages 1–8 (cited on page 20).
- [150] Tobias Nägeli et al. “Real-time Planning for Automated Multi-view Drone Cinematography”. In: *ACM Trans. Graph.* 36.4 (2017), 132:1–132:10. ISSN: 0730-0301. DOI: 10.1145/3072959.3073712. URL: <http://doi.acm.org/10.1145/3072959.3073712> (cited on page 146).
- [151] *NASA Task Load Index*. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000021488.pdf> (cited on page 72).
- [152] Marianna Obrist et al. “User Experience Research in the Semiconductor Factory: A Contradiction?” In: *Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part IV*. INTERACT’11. Lisbon, Portugal: Springer-Verlag, 2011, pages 144–151. ISBN: 978-3-642-23767-6. URL: <http://dl.acm.org/citation.cfm?id=2042283.2042298> (cited on page 147).
- [153] Paula M Oldfather, Allen S Ginsberg, and Harry M Markowitz. *Programming by questionnaire: How to construct a program generator*. Technical report. RAND CORP SANTA MONICA CALIF, 1966 (cited on page 108).
- [154] *OPC UA - Specification Part 2: Security Model. Specification*. 2018 (cited on page 21).
- [155] *OPC Unified Architecture – Part 1: Overview and concepts. Technical Report*. 2016 (cited on page 21).
- [156] *open62541: Open source implementation of OPC UA (OPC Unified Architecture) aka IEC 62541*. <https://github.com/open62541> (cited on page 22).
- [157] Gökhan Özbek, Fatih Kahraman, and Süleyman Baykut. “Robust video copy detection in large-scale TV streams using local features and CFAR based threshold”. In: *2016 IEEE International Conference on Digital Signal Processing (DSP)*. IEEE. 2016, pages 124–128 (cited on page 94).
- [158] Atakan Öztürk, Sinan Kayaligil, and Nur E Özdemirel. “Manufacturing lead time estimation using data mining”. In: *European Journal of Operational Research* 173.2 (2006), pages 683–700 (cited on page 26).
- [159] Hae-Sang Park and Chi-Hyuck Jun. “A simple and fast algorithm for K-medoids clustering”. In: *Expert systems with applications* 36.2 (2009), pages 3336–3341.
- [160] Dan Paulsson, Robert Gustavsson, and Carl-Fredrik Mandenius. “A soft sensor for bioprocess control based on sequential filtering of metabolic heat signals”. In: *Sensors* 14.10 (2014), pages 17864–17882.
- [161] Karl Pearson. “VII. Note on regression and inheritance in the case of two parents”. In: *proceedings of the royal society of London* 58.347-352 (1895), pages 240–242.

- [162] Ronald K Pearson. “Outliers in process modeling and identification”. In: *IEEE Transactions on control systems technology* 10.1 (2002), pages 55–63.
- [163] Catherine Plaisant. “The challenge of information visualization evaluation”. In: *Proceedings of the working conference on Advanced visual interfaces*. ACM. 2004, pages 109–116 (cited on page 74).
- [164] Klaus Pohl and Chris Rupp. *Basiswissen Requirements Engineering: Aus-und Weiterbildung nach IREB-Standard zum Certified Professional for Requirements Engineering Foundation Level*. dpunkt. verlag, 2015.
- [165] Sandy B Popovich, Shane R Chilton, and Bruce Kilgore. “Implementation of a test wafer inventory tracking system to increase efficiency in monitor wafer usage”. In: *1997 IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop ASMC 97 Proceedings*. IEEE. 1997, pages 440–443 (cited on page 91).
- [166] David Martin Powers. “Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation”. In: (2011).
- [167] Sebastian Rank et al. “Reducing simulation model complexity by using an adjustable base model for path-based automated material handling systems: a case study in the semiconductor industry”. In: *Proceedings of the 2015 Winter Simulation Conference*. IEEE Press. 2015, pages 2896–2907.
- [168] R RCoreTeam. *A language and environment for statistical computing*. <https://www.R-project.org>. Accessed on 28.01.2019. 2017 (cited on page 48).
- [169] Alexander Richter et al. “Der Mensch im Mittelpunkt der Fabrik von morgen”. In: *HMD Praxis der Wirtschaftsinformatik* 52.5 (2015), pages 690–712.
- [170] Alexander Richter et al. “Digital Work Design”. In: *Business & Information Systems Engineering* (2018). ISSN: 1867-0202. DOI: 10.1007/s12599-018-0534-4. URL: <https://doi.org/10.1007/s12599-018-0534-4>.
- [171] Frederic Ringsleben, Maik Benndorf, and Thomas Haenselmann. “Frequency-domain based video irregularity analysis”. In: *2017 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE. 2017, pages 572–575 (cited on page 96).
- [172] Frederic Ringsleben et al. “A New Approach using Characteristic Video Signals to Improve the Stability of Manufacturing Processes”. In: *2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE. 2018, pages 1–7 (cited on page 95).
- [173] Manfred Rosenberger and Alexander Stocker. “Eine Vorgehensweise zur Unterstützung der Einführung von Industrie-4.0-Technologien”. In: *Mensch und Computer 2017-Workshopband: Spielend einfach interagieren* 13 (2018), page 183.
- [174] Dan Rosso. *Global Semiconductor Sales Increase 13.7 Percent to \$ 468.8 Billion in 2018*. <https://www.semiconductors.org/global-semiconductor-sales-increase-13-7-percent-to-468-8-billion-in-2018/>. 2019.
- [175] Kosta Rozen and Néill M Byrne. “Using simulation to improve semiconductor factory cycle time by segregation of preventive maintenance activities”. In: *Proceedings of the 2016 Winter Simulation Conference*. IEEE Press. 2016, pages 2676–2684 (cited on page 107).
- [176] RStudio RStudioTeam. *Integrated development environment for R*. <https://www.rstudio.com/>. Accessed on 28.01.2019. 2016 (cited on page 48).
- [177] R Sadeghi, S Dauzère-Pérès, and C Yugma. “A multi-method simulation modelling for semiconductor manufacturing”. In: *IFAC-PapersOnLine* 49.12 (2016), pages 727–732 (cited on page 107).

- [178] Subhash C Sarin, Amrisha Varadarajan, and Lixin Wang. “A survey of dispatching rules for operational control in wafer fabrication”. In: *Production Planning and Control* 22.1 (2011), pages 4–24 (cited on page 86).
- [179] Anthony Scarsella and William Stofega. *Worldwide Smartphone Forecast Update, 2018–2022: September 2018*. <https://www.idc.com/getdoc.jsp?containerId=US44277718>. Retrieved: December 10, 2018, from [www.idc.com](http://www.idc.com). Sept. 2018 (cited on page 76).
- [180] Daniel Schel et al. “Manufacturing service bus: an implementation”. In: *11th CIRP Conf. Intell. Comput. Manuf. Eng.* Volume 67. 2017, page 6.
- [181] Robert Schmalzer et al. “Simulation based evaluation of different empty vehicle management strategies with considering future transport jobs”. In: *Proceedings of the 2017 Winter Simulation Conference*. IEEE Press. 2017, page 294 (cited on page 91).
- [182] Christoph Schmittner, Zhendong Ma, and Erwin Schoitsch. “Combined safety and security development lifecycle”. In: *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*. IEEE. 2015, pages 1408–1415 (cited on page 7).
- [183] Germar Schneider, Sophia Keil, and Gerhard Luhn. “Opportunities, challenges and use cases of digitization within the semiconductor industry”. In: *2018 29th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*. IEEE. 2018, pages 307–312 (cited on page 129).
- [184] Germar Schneider, Sophia Keil, and Gerhard Luhn. “Opportunities, Challenges and Use Cases of Digitization within the Semiconductor Industry”. In: *29th Annual SEMI Advanced Semiconductor Manufacturing Conference*. ASMC’18. 2018, pages 307–312. DOI: 10.1109/ASMC.2018.8373173.
- [185] S. Schneider et al. “Design and Development of a Benchmarking Testbed for the Factory of the Future”. In: *IEEE 20th Conference on Emerging Technologies Factory Automation (ETFA)*. IEEE. 2015, pages 1–7 (cited on page 17).
- [186] Wolfgang Scholl et al. “Simulation-enabled development lot journey smoothening in a fully-utilised semiconductor manufacturing line”. In: *2016 Winter Simulation Conference (WSC)*. IEEE. 2016, pages 2559–2567 (cited on pages 106, 107).
- [187] *Semi40*. <http://www.semi40.eu/>. Accessed on 25.01.2019 (cited on page 57).
- [188] *SEMICONDUCTORS - Fueling the Tech Industry’s Future as Never Before*. [https://www.accenture.com/t20180912T102159Z\\_\\_w\\_\\_/us-en/\\_acnmedia/PDF-85/Accenture-Semiconductor-Tech-Vision-2018-Executive-Summary.pdf](https://www.accenture.com/t20180912T102159Z__w__/us-en/_acnmedia/PDF-85/Accenture-Semiconductor-Tech-Vision-2018-Executive-Summary.pdf). 2018 (cited on page 126).
- [189] Enrico Senger and Hubert Österle. “Promet business engineering case studies (becs) version 2.0”. In: (2004).
- [190] *SFB 912: HIGHLY ADAPTIVE ENERGY-EFFICIENT COMPUTING*. <https://tu-dresden.de/ing/forschung/sonderforschungsbereiche/sfb912> (cited on page 127).
- [191] Colin Shearer. “The CRISP-DM model: the new blueprint for data mining”. In: *Journal of data warehousing* 5.4 (2000), pages 13–22 (cited on page 65).
- [192] A Sivakumar and R Gunasundari. “A Survey on Data Preprocessing Techniques for Bioinformatics and Web Usage Mining”. In: *International Journal of Pure and Applied Mathematics (IJPAM)* 117.20 (2017), pages 785–793 (cited on page 67).

- [193] *Smart Home*. [http://www.driverless-future.com/?page\\_id=384](http://www.driverless-future.com/?page_id=384). Retrieved: December 10, 2018, from [www.statista.com](http://www.statista.com). 2018 (cited on page 76).
- [194] Young Jun Son and Richard A Wysk. "Automatic simulation model generation for simulation-based, real-time shop floor control". In: *Computers in Industry* 45.3 (2001), pages 291–308 (cited on page 108).
- [195] Charles Spearman. "The proof and measurement of association between two things". In: *The American journal of psychology* 100.3/4 (1987), pages 441–471 (cited on page 72).
- [196] Michael Spitzer, Ibrahim Nanic, and Martin Ebner. "Distance Learning and Assistance Using Smart Glasses". In: *Education Sciences* 8.1 (2018). ISSN: 2227-7102. DOI: 10.3390/educsci8010021. URL: <http://www.mdpi.com/2227-7102/8/1/21>.
- [197] Darko Stanisavljevic and Michael Spitzer. "A Review of Related Work on Machine Learning in Semiconductor Manufacturing and Assembly Lines." In: *SAMI@ iKNOW*. 2016.
- [198] Darko Stanisavljevic et al. "Ein Industrie 4.0-Use Case in der Motorenproduktion". In: *Mensch und Computer 2018-Workshopband* (2018).
- [199] Susan Leigh Star and James R Griesemer. "Institutional ecology, translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39". In: *Social studies of science* 19.3 (1989), pages 387–420.
- [200] *Static and dynamic weaving*. [http://www.eclipse.org/eclipselink/documentation/2.7/concepts/app\\_dev005.htm](http://www.eclipse.org/eclipselink/documentation/2.7/concepts/app_dev005.htm). Accessed: 16-01-2019. 2019 (cited on page 116).
- [201] Alexander Stocker and Johannes Müller. "Exploring use and benefit of corporate social software: Measuring success in the Siemens case References+". In: *Journal of Systems and Information Technology* 18.3 (2016), pages 277–296.
- [202] Alexander Stocker and Klaus Tochtermann. *Wissenstransfer mit Wikis und Weblogs: Fallstudien zum erfolgreichen Einsatz von Web 2.0 in Unternehmen*. Springer-Verlag, 2011.
- [203] Alexander Stocker et al. "Exploring appropriation of enterprise wikis". In: *Computer Supported Cooperative Work (CSCW)* 21.2-3 (2012), pages 317–356.
- [204] Alexander Stocker et al. "Mensch-zentrierte IKT-Lösungen in einer Smart Factory". In: *e & i Elektrotechnik und Informationstechnik* 131.7 (2014), pages 207–211.
- [205] Alexander Stocker et al. "Datenbrillengestützte Checklisten in der Fahrzeugmontage". In: *Informatik-Spektrum* 40.3 (2017), pages 255–263.
- [206] Igor Stogniy and Wolfgang Scholl. "An automated-generated long-term production simulation model for semiconductor industry". In: *17th European advanced process control and manufacturing Conference*. 2017.
- [207] Igor Stogniy and Wolfgang Scholl. "The granularity of the automated-generated long-term simulation model". In: *18th European Advanced Process Control and Manufacturing Conference*. 2018.
- [208] B. Stoll et al. "Examining the Effects of Mobile Robotic Telepresence Systems in Collaborative Team Dynamics". In: *Workshop on Robots in Groups and Teams, Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW'17 Companion)*. ACM, 2017.
- [209] Nicole Stricker et al. "Reinforcement learning for adaptive order dispatching in the semiconductor industry". In: *CIRP Annals* 67.1 (2018), pages 511–514 (cited on page 48).

- [210] Masashi Sugiyama, Matthias Krauledat, and Klaus-Robert MÅžller. “Covariate shift adaptation by importance weighted cross validation”. In: *Journal of Machine Learning Research* 8.May (2007), pages 985–1005 (cited on page 82).
- [211] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [212] Derk J Swider. “A comparison of empirically based steady-state models for vapor-compression liquid chillers”. In: *Applied thermal engineering* 23.5 (2003), pages 539–556 (cited on page 82).
- [213] Gábor J Székely, Maria L Rizzo, Nail K Bakirov, et al. “Measuring and testing dependence by correlation of distances”. In: *The annals of statistics* 35.6 (2007), pages 2769–2794.
- [214] Markus Tauber and Christoph Schmittner. “Enabling Security and Safety Evaluation in Industry 4.0 Use Cases with Digital Twins”. In: *ERCIM News* (2018) (cited on page 8).
- [215] *The Feynman Lectures on Physics*. [http://www.feynmanlectures.caltech.edu/II\\_19.html](http://www.feynmanlectures.caltech.edu/II_19.html). 1990 (cited on pages 127, 128).
- [216] Frédéric Thiesse and Elgar Fleisch. “On the value of location information to lot scheduling in complex manufacturing processes”. In: *International Journal of Production Economics* 112.2 (2008), pages 532–547 (cited on page 86).
- [217] Divya Tomar and Sonali Agarwal. “A survey on pre-processing and post-processing techniques in data mining”. In: *International Journal of Database Theory & Application* 7.4 (2014), pages 99–128 (cited on page 67).
- [218] Katherine M. Tsui et al. “Exploring Use Cases for Telepresence Robots”. In: *Proceedings of the 6th International Conference on Human-robot Interaction*. HRI’11. Lausanne, Switzerland: ACM, 2011, pages 11–18. ISBN: 978-1-4503-0561-7. DOI: 10.1145/1957656.1957664. URL: <http://doi.acm.org/10.1145/1957656.1957664> (cited on page 146).
- [219] *Vaadin Framework*. <https://vaadin.com>. Accessed: 16-01-2019. 2019 (cited on page 115).
- [220] *Vaadin Framework Addon Directory*. <https://vaadin.com/directory>. Accessed: 16-01-2019. 2019 (cited on page 115).
- [221] Veerle Van den Eynden et al. “Managing and Sharing Data; a best practice guide for researchers”. In: (2011).
- [222] Birgit Vogel-Heuser, Thomas Bauernhansl, and Michael Ten Hompel. *Handbuch Industrie 4.0 Bd. 2*. Springer, 2017 (cited on page 30).
- [223] Siqi Wang et al. “Detecting Abnormality without Knowing Normality: A Two-stage Approach for Unsupervised Video Abnormal Event Detection”. In: *2018 ACM Multimedia Conference on Multimedia Conference*. ACM. 2018, pages 636–644 (cited on page 98).
- [224] Yi Wang et al. “Efficient test for nonlinear dependence of two continuous variables”. In: *BMC bioinformatics* 16.1 (2015), page 260.
- [225] Joe H Ward Jr. “Hierarchical grouping to optimize an objective function”. In: *Journal of the American statistical association* 58.301 (1963), pages 236–244.
- [226] Bernd Waschneck. “Unified Frontend and Backend Industrie 4.0 Roadmap for Semiconductor Manufacturing”. In: *International Conference on Knowledge Technologies and Data-Driven Business 2017 (i-KNOW 2017), Workshop Proceeding*. 2017.

- [227] Bernd Waschneck et al. "Production Scheduling in Complex Job Shops from an Industry 4.0 Perspective: A Review and Challenges in the Semiconductor Industry." In: *SAMI@iKNOW*. 2016, pages 1–12 (cited on page 38).
- [228] Hadley Wickham et al. "Tidy data". In: *Journal of Statistical Software* 59.10 (2014), pages 1–23 (cited on page 60).
- [229] HP Wiendahl. "Betriebsorganisation für Ingenieure: Mit 3 Tabellen. 8., überarb". In: *Aufl. München: Hanser* (2014) (cited on page 36).
- [230] Muh-Cherng Wu, CS Chien, and KS Lu. "Planning yields in recycling test wafers". In: *2004 Semiconductor Manufacturing Technology Workshop Proceedings (IEEE Cat. No. 04EX846)*. IEEE. 2004, pages 141–144 (cited on page 91).
- [231] Francis WH Yik and Vincent KC Lam. "Chiller models for plant design studies". In: *Building Services Engineering Research and Technology* 19.4 (1998), pages 233–241 (cited on page 82).
- [232] Robert K Yin. *Applications of case study research*. sage, 2011.
- [233] Jess Macy Yu. *Chipmaker Nvidia's CEO sees fully autonomous cars within 4 years*. <https://www.reuters.com/article/us-nvidia-ai-chips/chipmaker-nvidias-ceo-sees-fully-autonomous-cars-within-4-years-idUSKBN1CV192?feedType=RSS&feedName=technologyNews>. Retrieved: December 10, 2018. Oct. 2017 (cited on page 76).
- [234] Shi Zhong and Joydeep Ghosh. "Scalable, balanced model-based clustering". In: *Proceedings of the 2003 SIAM International Conference on Data Mining*. SIAM. 2003, pages 71–82.
- [235] Keliang Zhou, Taigang Liu, and Lifeng Zhou. "Industry 4.0: Towards future industrial opportunities and challenges". In: *2015 12th International conference on fuzzy systems and knowledge discovery (FSKD)*. IEEE. 2015, pages 2147–2152.