

RWTH THEMEN

Forschungsmagazin

2/2020

Profilbereich
„Information & Communication Technology“

Inhalt

- 4 **Der Profilbereich
„Information & Communication Technology“**
von Joost-Pieter Katoen
- 8 **Planoptimierung für Roboterteams
in der Produktionslogistik**
von Erika Ábrahám, Gerhard Lakemeyer
- 12 **Reden virtuelle Menschen künstlich? Nein, natürlich!**
Audiovisuelle virtuelle Welten made in Aachen
von Andrea Bönsch, Jonathan Ehret, Janina Fels,
Torsten Wolfgang Kuhlen, Michael Vorländer
- 18 **Mit „Digitalen Schatten“ Daten verdichten und darstellen**
Der Exzellenzcluster „Internet der Produktion“ forscht
über die Produktionstechnik hinaus
von Matthias Jarke, Wil van der Aalst, Christian Brecher,
Matthias Brockmann, István Koren, Gerhard Lakemeyer,
Bernhard Rumpe, Günther Schuh, Klaus Wehrle, Martina Ziefle
- 25 **Materie und Licht für Quanteninformation**
Forschung mit regionalen Partnern
von Marian Barsoum, Hendrik Bluhm
- 28 **Das interdisziplinäre Graduiertenkolleg
„UnRAVeL – Uncertainty and Randomness
in Algorithms, Verification and Logic“**
Forschung zur Eisenbahninfrastruktur und dem Eisenbahnbetrieb
von Joost-Pieter Katoen, Nils Nießen, Stephan Zieger
- 34 **Responsible Data Science**
Mit Datenwissenschaften Entscheidungsprozesse gestalten
von Saskia Nagel, Markus Strohmaier
- 38 **Die Mathematik der Entscheidungen**
Mit Operations Research zu bestmöglichen
Handlungsempfehlungen
von Marco Lübbecke
- 44 **Visuelles Szenenverstehen in dynamischen Umgebungen**
Deep Learning für Computer-Vision-Anwendungen
von Bastian Leibe
- 50 **Sprachverarbeitung und künstliche Intelligenz**
von Hermann Ney
- 54 **Fehler in der Steuerungssoftware –
eine wissenschaftliche Herausforderung**
Technische Systeme werden sicherer mit Informatik-Methoden
von Stefan Kowalewski
- 58 **Alle Wege durch das Labyrinth**
Die automatisierte Suche nach Softwarefehlern
von Daniel Schemmel, Julian Büning, Felix Rath, Klaus Wehrle
- 63 **Energieeffiziente KI**
Neuromorphe Hardware für künstliche intelligente Systeme
von Tobias Gemmeke, Susanne Hoffmann-Eifert,
Alexander Krüger, Max Lemme, Rainer Waser, Dirk Wouters
- 70 **Vom Hardware-Trojaner zum Blackout**
Neue Cybersecurity-Methoden für Energienetze
und Mikroprozessoren
von Rainer Leupers, Antonello Monti
- 78 **Von schwenkenden Beams und hohen Datenraten**
Der 5G-Rollout aus Strahlenschutzsicht
von Dirk Heberling, Thomas Kopacz, Sascha Schießl,
Anna-Malin Schiffarth
- 82 **Quantencomputing**
Ein Paradigmenwechsel in der Datenverarbeitung
von Marian Barsoum, Hendrik Bluhm
- 86 **Namen & Nachrichten**

Der Profilbereich „Information & Communication Technology“

Informations- und Kommunikationstechnologien (IKT) sind Bestandteil des täglichen Lebens und kommen als Schlüsseltechnologie des 21. Jahrhunderts in fast allen Forschungsbereichen vor. Der Profilbereich „Information & Communication Technology“, kurz ICT, ist einer von acht Profilbereichen der RWTH Aachen. Diese Profilbereiche bündeln die wissenschaftliche Expertise zu den Forschungsschwerpunkten der RWTH, sie sind Bestandteil der Forschungslandschaft und dienen als Kommunikationskanäle innerhalb der Hochschule.

Im Profilbereich ICT steht die Informationstechnologie im Vordergrund, also die Anwendung von Computern, eingebetteten und steuernden Geräten sowie menschlicher und computergestützter Informationsverarbeitung, um Informationen in digitaler Form zu speichern, abzurufen, zu manipulieren, darzustellen und zu verstehen. Dabei wird stets die Kommunikationstechnologie im weiteren Sinne genutzt.

Mit zahlreichen Maßnahmen und Initiativen fördert der Profilbereich Forschung rund um die IKT. So wurden Workshops, eine Referenzenreihe und „Researcher Spotlights“ sowie der Young Researcher Award ins Leben gerufen, mit dem das innovative Forschungspotenzial der Studierenden und jungen Wissenschaftlerinnen und Wissenschaftler gefördert werden soll. Vor einigen Jahren initiierte der Profilbereich das Projekthaus „ICT – Grundlagen einer digitalisierten Industrie, Wirtschaft und Gesellschaft“. Ziel war, interdisziplinäre Forschungsprojekte für die laufende digitale Transformation zu fördern. Eine Anschubfi-

nanzierung für 20 Projekte sowie eine Plattform für die Interaktion zwischen den Forschenden wurden ermöglicht. Hieraus sind in der Folge auch Projekte in Zusammenarbeit mit einer Vielzahl externer Partner entstanden.

Die Forschung innerhalb des Profilbereichs ICT genießt international hohes Ansehen. Jenes bestätigen mehrere ERC-Grants, mehrere Alexander von Humboldt-Professoren und eine starke Beteiligung an EU-Projekten. Der interdisziplinäre Charakter steht dabei im Fokus, welches sich anhand einer intensiven Beteiligung an Graduiertenkollegs und Verbundprojekten mit diversen Forschungsbereichen beweisen lässt. Es bestehen Kooperationen mit fast allen Fakultäten und allen anderen Profilbereichen der RWTH. Der Profilbereich ICT ist an zwei Exzellenzclustern der RWTH beteiligt: an „ML4Q – Matter and Light for Quantum Computing“ und „Internet der Produktion“.

Diese Ausgabe der RWTH THEMEN gibt einen Einblick in Forschungsprojekte aus fünf Forschungsbereichen:

- Artificial Intelligence (AI),
- Data Science,
- Dependable Digitalization,
- Next Generation Computing and Communication Platforms und
- Quantum Computing.

Diese Forschungsgebiete knüpfen stark aneinander an: AI und Data Science teilen diverse Aspekte, gleiches gilt für Quantum Computing und Next Generation Computing

and Communication Platforms. Diese vier Bereiche können als horizontal organisiert betrachtet werden. Das Schwerpunktgebiet Dependable Digitalization hat einen allumfassenden Charakter. Es hat Berührungspunkte zu allen vier Forschungsbereichen und kann als vertikale Dimension der anderen Gebiete angesehen werden. Diese Verbindungen reichen von sicherer und erklärbarer AI über vertrauenswürdige Datenwissenschaften zu verlässlichen und geschützten Berechnungen und Kommunikationsmitteln.

Artificial Intelligence

Artificial Intelligence, kurz AI oder KI, hat das Potenzial, ganze Industrien zum Erliegen zu bringen und zu revolutionieren. Etliche Wellen des Fortschritts sorgten für die Grundlage und Erfolge neuronaler Netzwerke, regelbasierter KI-Systeme und Techniken des statistischen maschinellen Lernens. Die heutige KI unterscheidet sich stark von der KI vor ein oder zwei Jahrzehnten. Insbesondere Entwicklungen in Deep Learning (und hier in der Mustererkennung) haben die Forschung erheblich vorangebracht, sodass sie nun in den verschiedensten Bereichen mit enormer wirtschaftlicher Wirkung Anwendung findet. An der RWTH werden KI-Methoden wie beispielsweise probabilistische Algorithmen für die Navigation und (Deep) Learning zum Erkennen von Objekten (Computer Vision) angewendet, welche für die Entwicklung von Robotern in der Produktionslogistik eingesetzt werden.

Data Science

Data Science verändert die Rolle der IKT in der Gesellschaft und allen Wirtschaftssektoren rapide. Sie ist ein interdisziplinäres Feld mit der Zielsetzung, Daten in realen Wert umzuwandeln. Daten können strukturiert oder unstrukturiert sein, groß oder klein, statisch oder fließend. Wert wiederum kann in Form von Vorhersagen, automatisierten Entscheidungen, Datenmodellen oder jeder Art von Datenvisualisierung, die Erkenntnisse liefert, bereitgestellt werden. Data Science umfasst die Extraktion, Vorbereitung, Exploration, Transformation, Speicherung und Abfrage von Daten, verschiedene Arten von Data Mining und Learning, Recheninfrastrukturen, die Präsentation von Erklärungen und Vorhersagen sowie die Nutzung von Ergebnissen unter Berücksichtigung ethischer, sozialer, rechtlicher und geschäftlicher Aspekte. Viele Sektoren werden einen grundlegenden Wandel durchlaufen, sodass in den kommenden Jahren viele Arbeitsplätze durch Software oder Hardware-Roboter ersetzt werden. Der wichtigste limitierende Faktor in dieser Transformation werden Data-Science-Fähigkeiten sein.

Dependable Digitalization

Millionen von Programmzeilen sind in modernen Autos, medizinischen Robotern, Luft- und Raumfahrtssystemen und kritischen Infrastrukturen vorhanden. Der Einfluss von Software nimmt mit dem Bedürfnis nach mehr Autonomie rasch zu. Es versteht sich von selbst, dass diese sicherheitskritischen Systeme stark von korrekt funktionierender

Software abhängen und auf zuverlässigen und sicheren Recheninfrastrukturen laufen müssen. Trotz der Tatsache, dass IKT sich überall und auf jeden von uns auswirken, wird der zu erbringende Aufwand, um sowohl Software als auch Hardware zuverlässig, wartbar, sicher und über längere Zeiträume nutzbar zu machen, häufig unterschätzt. Teure und verheerende Softwarefehler sowie gehackte IKT-Systeme sind jedoch regelmäßig Schlagzeilen in den Nachrichten. Der Forschungsbereich Dependable Digitalization befasst sich damit, Software und IKT-Systeme zuverlässig und sicher zu gestalten.

Next Generation Computing and Communication Platforms

In Zukunft wird die Interaktion mit Computern durch virtuelle und erweiterte Realität so verbessert werden, dass sich die Interaktion wie eine Mensch-zu-Mensch-Interaktion anfühlt. All dies stützt sich zum Teil auf den Einsatz von Techniken, die auf KI basieren. Es wird erwartet, dass Systeme ihre physische Umgebung nicht nur passiv beobachten, sondern auch kontinuierlich mit der physischen Welt interagieren, um diese aktiv zu steuern. Paradebeispiele sind fortgeschrittene Fahrerassistenzsysteme, halbautonomes Fahren und medizinische KI. Diese Entwicklungen stellen uns vor große Herausforderungen: Menschen müssen diesen Systemen vertrauen können, das heißt ihre Kontrolle über die physische Welt muss sich innerhalb sicherer Grenzen bewegen, keinen Schaden anrichten und Daten gegebenenfalls geheimhalten. Geräte mit integrierten Schaltkreisen benö-

tigen zudem viel Energie. Um die Datenflut aus dem anstehenden Internet der Dinge zu stemmen, ist eine intelligente lokale Datenverarbeitung auf entfernten Geräten erforderlich, die den Energieverbrauch minimiert. Dazu gehört die Abkehr von traditionellen von Neumann-Architekturen und ein Umdenken in der Geräte- und Kommunikationstechnik.

Quantum Computing

Quantum Computing ist ein Thema, welches die Erwartung weckt, Hochleistungs-Computing in Anwendungen wie Optimierung, Quantenchemie für Material-, Katalysator- und Medikamentenentwicklung, maschinellem Lernen und einigen anderen Bereichen zu revolutionieren. Die Idee besteht darin, Quantenzustände mit ihren besonderen Eigenschaften wie Überlagerung und Verschränkung als Grundlage für die Informationsverarbeitung zu nutzen und so eine exponentielle Parallelität und einen Speichereuwachs für ausgewählte Anwendungen zu erreichen. Das Feld hat ein enormes Potenzial. Spannende Herausforderungen sind, wie man solche Quantumcomputer effizient programmieren kann, und wie man ein Betriebssystem gestalten kann.

Autor

Univ.-Prof. Dr. ir. Dr. h.c. Joost-Pieter Katoen ist Inhaber des Lehrstuhls für Informatik 2 (Softwaremodellierung und Verifikation) und Sprecher des Profilbereichs „Information & Communication Technology“.

The Profile Area “Information & Communication Technology”

Information and Communication Technology (ICT) is an integral part of everyday life and, as a key technology in the 21st century, it has connections to almost all research fields. The Profile Area Information & Communication Technology (ICT) is one of eight Profile Areas of RWTH Aachen University. The Profile Areas, which are pivotal organizational parts of RWTH's research landscape, pool scientific expertise in the university's key research areas and serve as communication channels within it.

The ICT Profile Areas is concerned with information technology, that is, the application of computers, embedded and controlling devices and human- as well as computer-assisted information processing to store, retrieve, manipulate, present and understand information in digital form, while using communication technology in a broader sense. To this end, the ICT Profile Area has initiated several activities. In order to reach out to members from RWTH, it has created workshops, a speaker series, “Researcher Spotlights”, and the Young Researcher Award, which aims to recognize and support the innovative research potential of the excellent students and early-career researchers at RWTH. A few years ago, we initiated the project house “ICT Foundations of a Digitized Industry, Economy and Society”. Its aim was to foster interdisciplinary joint research projects of RWTH groups on ICT for the ongoing digital transformation. The project house provided seed funding for 20 projects and a platform for interaction between the participating researchers. This has led to various

follow-up projects in cooperation with a large number of external partners.

The research within the ICT Profile Area has an internationally renowned reputation (as witnessed by multiple ERC grants, several Alexander-von-Humboldt professors, and a strong participation in European projects) and is characterized by a strong interdisciplinary character. This is e.g. witnessed by an intense participation in Collaborative Research Centers, Research Training Groups, FET Flagship Projects, Priority Programs, and so forth. The Profile Area enjoys collaborations with almost all faculties and all other profile areas at RWTH. It participates in two out three Clusters of Excellence at RWTH, namely “Material and Light for Quantum Computing (ML4Q)” and the “Internet of Production” (IoP).

This issue of RWTH THEMEN provides insight in active research projects in five prominent research areas:

- Artificial Intelligence,
- Data Science,
- Dependable Digitalization,
- Next Generation Computing and Communication Platforms, and
- Quantum Computing.

There are strong connections between these research areas. AI and Data Science have various aspects in common, and the same holds for Quantum Computing and Next Generation Computing and Communication Platforms. These four areas can be considered to be horizontally organized. The Dependable Digitalization research area is of

a more overarching, comprehensive character: it is relevant for the other four areas and can be seen as a vertical dimension of them. These connections range from secure and explainable AI through trustworthy data science to reliable and secure software and communication protocols.

Artificial Intelligence

Artificial Intelligence (AI) has the potential to disrupt and revolutionize entire industries. Different waves of progress brought the foundations and successes of neural networks, rule-based AI systems, and statistical machine learning techniques. The developments in Deep Learning have significantly advanced the field, in particular pattern recognition, and has reached out to various application areas with a potential huge economic impact. At RWTH, AI methods are e.g. used in the development of robots in production logistics, such as probabilistic algorithms to navigate and (deep) learning to detect objects (computer vision).

Data Science

Data Science is rapidly transforming the role of ICT in society and in all sectors of economy. It aims to turn data into real value. Data may be structured or unstructured, big or small, static or streaming. Value may be provided in the form of predictions, automated decisions, models learned from data, or any type of data visualization delivering insights. Data Science includes data extraction, data preparation, data exploration, data transformation, storage and retrieval, computing

infrastructures, various types of mining and learning, presentation of explanations and predictions, and the exploitation of results taking into account ethical, social, legal, and business aspects. It is foreseen that many sectors will undergo fundamental change, with many jobs being replaced by software or by robots in the coming years. The main limiting factor in this transformation process will be data science skills.

Dependable Digitalization

Millions of lines of program code are present in, e.g., modern cars, medical robots, aerospace systems and critical infrastructures. The impact of software is rapidly increasing with the need for more autonomy. It goes without saying that these safety-critical systems all heavily rely on the correct functioning of the software and need to run on reliable and secure computing infrastructures. Despite the fact that ICT is ubiquitous, the effort required to make both software and hardware reliable, maintainable, secure and usable over a long period of time is often underestimated. However, expensive and disastrous software errors and hacked ICT systems continue to make headlines in the news. The research area Dependable Digitalization is all about making software, data, and ICT systems reliable and secure.

Next Generation Computers and Communication

In the future, the interactions with computers will be enhanced by virtual and augmented reality – partially based on AI techniques

– such that they feel like human-to-human interaction. Systems will not only passively observe the physical environment, but continuously interact with it to actively control it. Prime examples are advanced driver assistance systems, semi-autonomous driving, and medical AI. These developments pose significant challenges. Users must be able to trust these systems, which means that their control over the physical world must be secure and take place within clearly defined boundaries, must not cause any damage and, where appropriate, must keep data secret. Furthermore, computing devices need an enormous amount of energy. In order to stem the flood of data from the Internet of Things, intelligent local data processing on remote devices is required, minimizing the energy consumption. This includes breaking away from traditional Von Neumann architectures and rethinking device and communication technology.

Quantum Computing

Quantum Computing is a hot topic raising expectations to revolutionize high performance computing in high impact applications like optimization, quantum chemistry for material, catalyst and drug design, machine learning, and several others. The key idea is to use quantum states with their peculiar properties like superposition and entanglement as a basis for information processing, thus achieving an exponential parallelism and memory boost for select applications. The field has enormous potential, and exciting challenges are how to program such quan-

tum computers efficiently and how to design effective operating systems for them.

Author

Univ.-Prof. Dr. ir. Dr. h.c. Joost-Pieter Katoen is holder of the Chair for Software Modeling and Verification (Computer Science 2) and spokesperson for the Information & Communication Technology Profile Area at RWTH Aachen University.

Planoptimierung für Roboterteams in der Produktionslogistik

One of the visions of Industry 4.0 is the flexible production of goods with the help of robots and machines. To facilitate research in this area in a competitive setting, the RoboCup Federation founded the Logistics League, where teams of robots compete in the production of goods with high variability. The Carologistics Team of RWTH Aachen University, in cooperation with FH Aachen University of Applied Sciences, has successfully participated in this competition for many years. One of the research challenges in this domain is to come up with optimal plans for the robots to fulfill given product orders. In this project we show how formal methods from computer science, in particular, satisfiability modulo theories, can be successfully employed to solve this problem.



Bild 1: Das Aachener Team „Carologistics“ gewinnt das Finale der RoboCup Logistics League in Hefei, China.

Eine der Visionen von Industrie 4.0 betrifft die Flexibilisierung von Produktionsprozessen. So sollen variantenreiche Produkte selbst in kleinen Losgrößen profitabel herzustellen sein. Unter anderem erfordert dies eine modulare Fertigung, idealerweise unterstützt von autonomen Robotern, die mit Menschen und Maschinen interagieren. Um dieser Vision näherzukommen, wurde vor circa zehn Jahren von der weltweit agierenden Robotikorganisation „RoboCup“ die Logistics League ins Leben gerufen. Dabei handelt es sich um einen spielerischen Wettbewerb, in dem jeweils zwei Teams autonomer Roboter mit Hilfe von Maschinen Produkte mit hoher Variabilität herstellen. Die Produktvarianten und Deadlines für die Fertigstellung werden erst während des Wettbewerbs den Robotern mitgeteilt, siehe Bild 1. Das Team „Carolo-

gistics“ von der RWTH Aachen und der FH Aachen nimmt regelmäßig an diesem internationalen Wettbewerb teil und wurde 2019 bereits zum fünften Mal Weltmeister. Neben der Bewältigung robotertypischer Herausforderungen wie Navigation und Manipulation von Objekten sind verteilte Handlungspläne für die jeweils vorliegenden Produktionsaufträge zu erstellen. Diese müssen die Roboter effizient und unter Einhaltung der Deadlines erfüllen. Das Lehr- und Forschungsgebiet Informatik 5 (Wissensbasierte Systeme) und das Lehr- und Forschungsgebiet Theorie hybrider Systeme haben hierzu ein Projekt durchgeführt.

Die RoboCup Logistics League

Bei diesem Wettbewerb stehen sich zwei Teams mit jeweils drei Robotern auf einem 14 x 8 m² großen Spielfeld gegenüber. Die Roboter eines Teams können untereinander sowie mit der sogenannten Refbox, die den Spielablauf kontrolliert, kommunizieren. Jedes Team verfügt über sechs Maschinen verschiedenen Typs. Die Maschinen werden vor jedem Spiel zufällig auf dem Spielfeld verteilt, zunächst müssen die Roboter diese suchen und identifizieren. Zwar hat jedes Team seine Maschinen, wegen der zufälligen Verteilung auf dem Spielfeld kommt es aber immer wieder zu Begegnungen und damit Behinderungen der Roboter. Während der fünfzehnminütigen Produktionsphase gilt es zufällig von der Refbox erzeugte Aufträge zu erledigen. Jeder Auftrag spezifiziert ein Produkt und ein Zeitfenster. Produkte bestehen aus einer farbigen zylinderförmigen Basis, bis zu drei farbigen Ringen und einem Deckel, so sind fast 250 Varianten möglich. Zunächst muss die Basis von einer Base Station (BS) geholt und dann an zwei Ring Stations (RS 1+2) mit Ringen versehen wer-

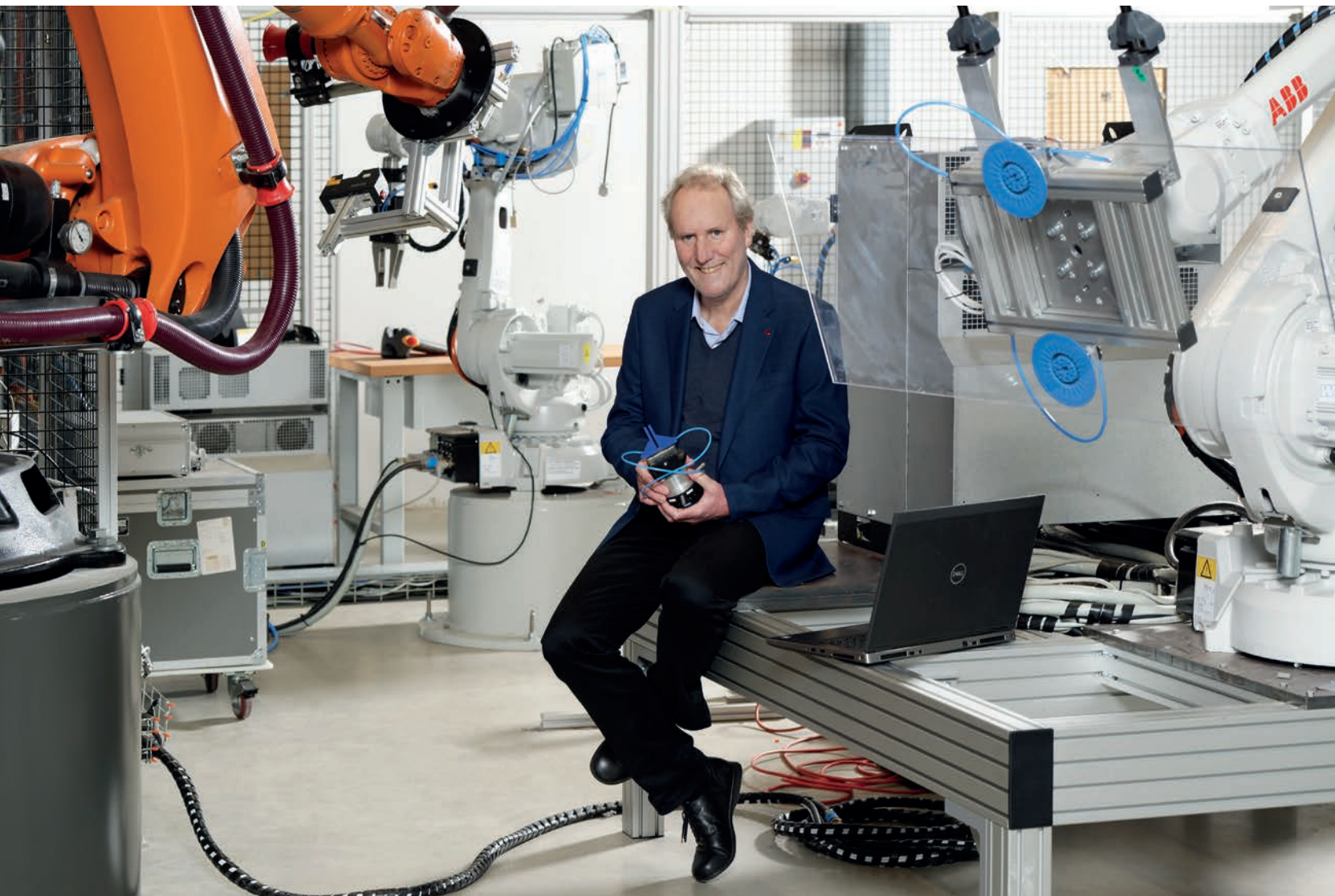


Bild 2: Das Lehr- und Forschungsgebiet Informatik 5 (Wissensbasierte Systeme) arbeitet unter Leitung von Professor Gerhard Lakemeyer an den Grundlagen der Wissensrepräsentation mit Anwendungen in der kognitiven Robotik.
Foto: Peter Winandy

den, wobei das Aufsetzen der Ringe zusätzliches Material benötigen kann. Abschließend wird an der Cap Station (CS) der Deckel montiert und das fertige Produkt abgeliefert. Für abgelieferte Produkte werden, je nach Komplexität, Punkte vergeben. Ziel des Spiels ist es, möglichst viele Punkte zu sammeln. Die Hauptaufgabe der Roboter besteht darin, Material oder Teilprodukte zu den Maschinen zu transportieren beziehungsweise von dort abzuholen. Während die Produktionsabläufe für jedes Produkt im Wesentlichen vordefi-

nier sind, liegt eine Herausforderung darin, die Roboter optimal einzusetzen, um entweder gleichzeitig an demselben Produkt oder parallel an verschiedenen Aufträgen zu arbeiten. Erschwerend kommt hinzu, dass Ausführungsfehler passieren, etwa durch Fallenlassen eines Teilprodukts, auch können Roboter und Maschinen ausfallen, sodass eine Neuplanung notwendig wird. Neben den Robotern und Maschinen wurde am Lehr- und Forschungsgebiet Informatik 5 (Wissensbasierte Systeme) eine Simulations-



Bild 3: Schematischer Produktionsablauf zur Herstellung eines C3-Produkts, bestehend aus einer Basis, drei Ringen und einem Deckel.



Bild 4: Das Lehr- und Forschungsgebiet Theorie hybrider Systeme entwickelt unter Leitung von Professorin Erika Ábrahám automatisierte Methoden zur effektiven Lösung komplexer kombinatorischer Probleme.
Foto: Peter Winandy

umgebung entwickelt, siehe Bild 5, die es erlaubt, Spiele physikalisch realistisch zu simulieren. Dabei wird die Software genutzt, die auch auf den echten Robotern betrieben wird. So können Softwarekomponenten wie etwa Planungsverfahren entwickelt und getestet werden, ohne mit den realen Robotern zu arbeiten.

Planungsstrategien mit Satisfiability Checking

Viele Probleme können mithilfe logischer Formeln kodiert werden. Solche Kodierungen werden zum Beispiel eingesetzt, um zu überprüfen ob ein System erwartungsgemäß funktioniert, um Systemabläufe mit bestimmten Eigenschaften zu identifizieren, um Herstellungsprozesse zu optimieren oder optimale Logistikpläne zu erstellen. Durch eine Kodierung erhält man eine präzise Problembeschreibung. Zudem gibt es Programme, die logische Probleme automatisch und effizient lösen. Auf dieser Basis kann man anschließend die Lösung für das ursprüngliche Problem „de-codieren“.

„Satisfiability Checking“ ist ein junges Forschungsgebiet in der Informatik. Ziel ist es, Methoden und Programme zu entwickeln, die in der Lage sind, logische Probleme mit Hilfe von Computern automatisch zu lösen. Die verwendeten Algorithmen basieren typischerweise auf mathematischer Logik und können je nach Problemklasse sehr komplex sein. Intensive Forschung, eine aktive Community und regelmäßige Wettbewerbe haben dazu beigetragen, dass die Entwicklung mit großen Schritten voranschreitet.

SAT Solver sind eine Klasse dieser Programme für aussagenlogische Probleme, also für Formeln, in denen Variablen mit möglichen Werten wahr oder falsch logisch mit den Operatoren „und“, „oder“, „nicht“ und so weiter kombiniert werden. Moderne SAT Solver sind in der Lage, extrem große aussagenlogische Probleme in kurzer Zeit lösen zu können. Diese Fähigkeit macht sie nicht nur akademisch beliebt, sondern führt auch zum häufigeren industriellen Einsatz. Typische Einsatzgebiete sind kombinatorischer Natur, zum Beispiel können SAT Solver solche Eingaben bestimmen, für die fehlerhafte Chips unerwünschte Ausgaben generieren.

Man kann die Aussagekraft erhöhen und die logische Kodierung von praktischen Problemen vereinfachen, indem man in den Formeln statt aussagenlogischer Variablen zum Beispiel Gleichungen und Ungleichungen zwischen arithmetischen Termen erlaubt.

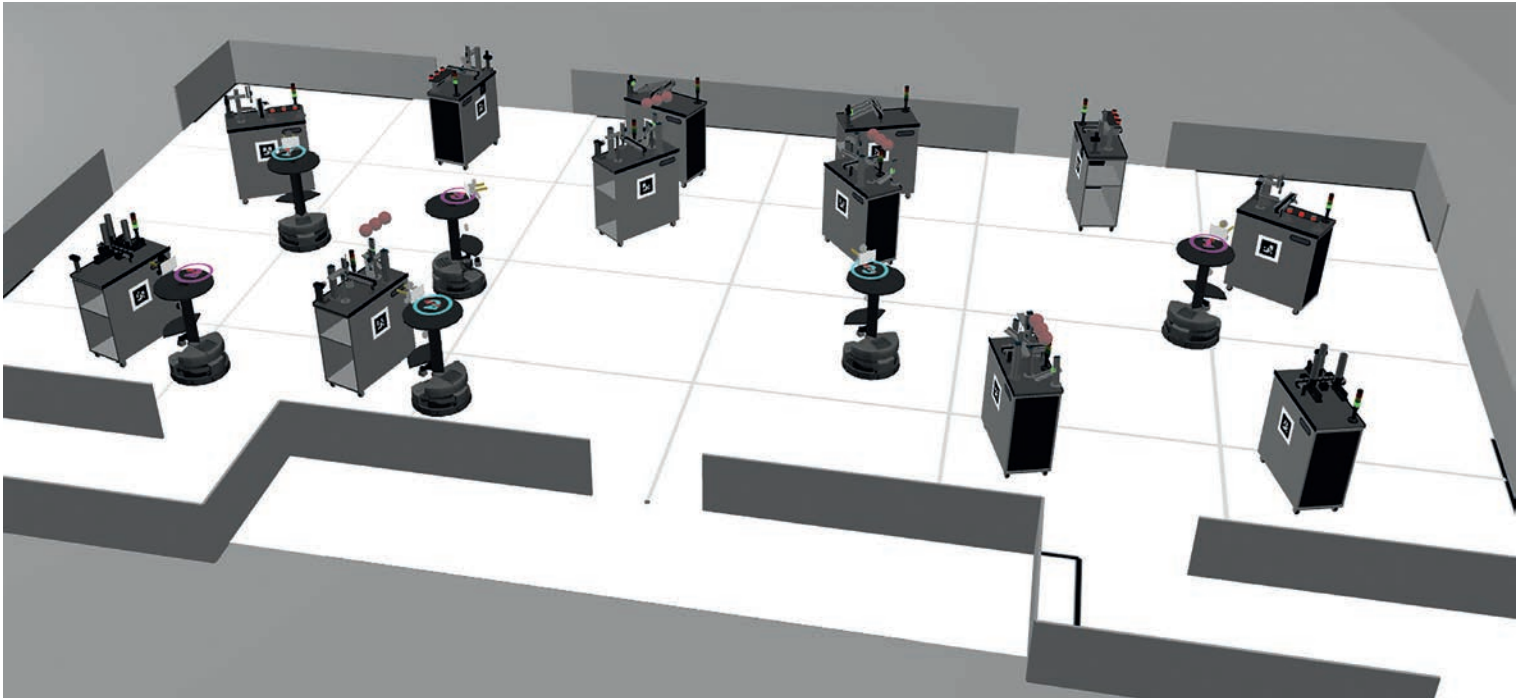


Bild 5: Eine SMT-Solving-basierte Methode konnte 2018 im Wettbewerb „Planning and Execution Competition for Logistics Robots in Simulation“ traditionelle heuristische Planning-Ansätze überbieten.

In arithmetischen Theorien sind es zum Beispiel Gleichungen und Ungleichungen zwischen arithmetischen Termen. Programme, die die Erfüllbarkeit von solchen Formeln automatisch mit Satisfiability Checking Methoden bestimmen können, nennt man SAT-Modulo-Theorien (SMT) Solver. Auch SMT Solver erfreuen sich wachsender Beliebtheit und einer zunehmenden Vielfalt von Einsatzbereichen. Die Theorien erlauben zum Beispiel die Verwendung von physikalischen Größen wie Zeit oder Geschwindigkeit in der Problembeschreibung. Das Lehr- und For-

schungsgebiet Theorie hybrider Systeme hat in diesen Entwicklungen eine international führende Rolle, sowohl in der Technologieentwicklung als auch in der interdisziplinären Stärkung und beim Technologietransfer in die Industrie. Die Planung der strategischen Handlungen für Roboter hat sowohl kombinatorische als auch temporale und arithmetische Aspekte. Daher liegt der Einsatz von SMT Solvern in diesem Bereich auf der Hand, allerdings waren solche Bemühungen nur mäßig erfolgreich. Dies liegt teilweise daran, dass

SMT Solver zwar „off-the-shelf“ verwendet werden können, für einen effektiven Einsatz bei komplexen Problemen ist es von Vorteil, wenn man die Arbeitsweise der Solver in der Kodierung des Problems berücksichtigt, sodass die entstehenden Formeln für den Solver einfacher zu lösen sind. Im Projekt wurde hier angesetzt und in enger Zusammenarbeit der beiden Lehr- und Forschungsgebiete eine neuwertige Kodierungstechnik entwickelt. Diese konnte den Einsatz von SMT Solvern so weit beschleunigen, dass eine Anwendung in Echtzeit möglich wurde. Zwar steht ein Test dieses Ansatzes mit Robotern in einem Wettbewerb noch aus, im internationalen Wettbewerb „Planning and Execution Competition for Logistics Robots in Simulation“ konnte jedoch mit dem SMT-basierten Ansatz 2018 der erste Platz erreicht und damit traditionelle heuristische Ansätze überboten werden.

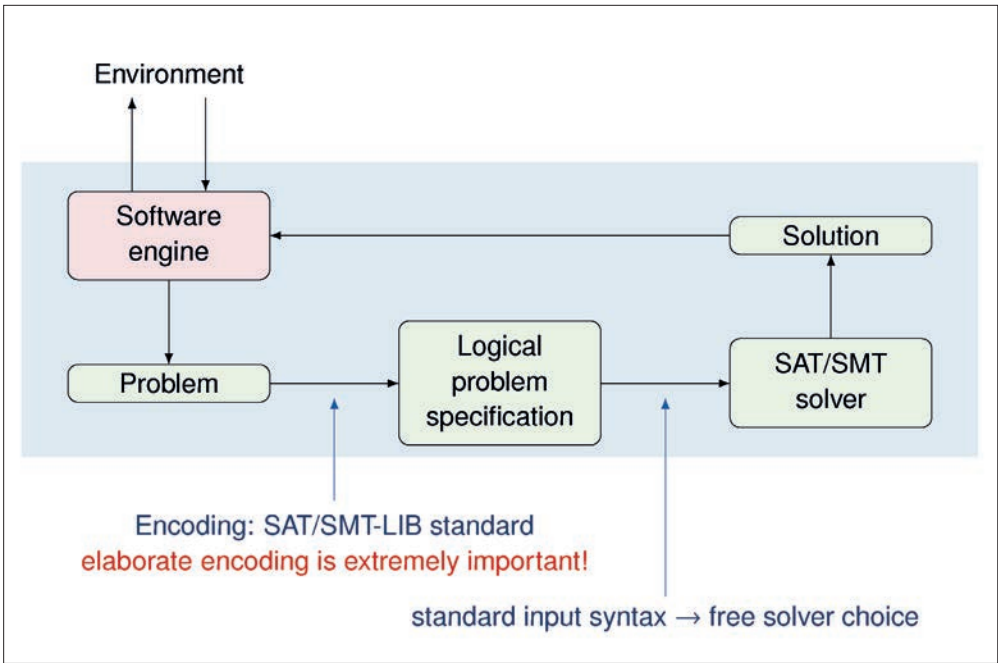


Bild 6: Mit Satisfiability-Modulo-Theorien (SMT) Solver können logisch kodierte Fragestellungen automatisiert gelöst werden. Erfolgreiche Kooperationen mit industriellen Partnern zeigen beeindruckendes Potenzial.

Autoren
 Univ.-Prof. Dr. Erika Ábrahám betreut das Lehr- und Forschungsgebiet Theorie hybrider Systeme.
 Univ.-Prof. Gerhard Lakemeyer, Ph.D., betreut das Lehr- und Forschungsgebiet Informatik 5 (Wissensbasierte Systeme).

Reden virtuelle Menschen künstlich? Nein, natürlich!

Audiovisuelle virtuelle Welten made in Aachen

Designing computer-controlled, human-like virtual agents (VAs) for highly dynamic and lifelike virtual worlds is a challenging task. To convincingly represent conversational interaction partners, VAs need to plausibly mimic visual and acoustic human behavior. This includes lip, eye, and body movements, as well as speech and its sound propagation – all in real-time and at a high-quality level. Combining their expertise in Virtual Acoustics and Social Virtual Reality, researchers from RWTH Aachen University are meeting this challenge. Beyond enriching the VA's design with, inter alia, co-verbal gestures, they are improving the auralization of the spoken word by means of a realistic, physically-based acoustic simulation. Directional voice radiation allows listeners to relate the sound not only to the speaker's location in the virtual world, but also to its orientation, e.g., whether they are located face-to-face or face-to-back. To maximize the realism of the acoustic reproduction, dynamic directivity is applied as well, taking the phoneme-dependent characteristics of the spoken word into account.

The resulting audiovisual virtual worlds allow the human sciences to systematically investigate the influence of spatial acoustics and visual stimuli on human cognition. To further improve the understanding in the area of auditory cognition, RWTH initiated the DFG Priority Program AUDICTIVE with colleagues from the universities of Ilmenau and Oldenburg.

Seit dem ersten Hype in den 90er Jahren ist die Virtuelle Realität (VR) auf dem Weg zu einem seriösen und leistungsstarken Werkzeug für wissenschaftliche und industrielle Anwendungen. Mit den neuesten Entwicklungen im Bereich der Hardware- und Software-technologien wurde nun ein hohes Maß an Realitätsnähe erreicht. So lassen sich selbst hochdynamische, „lebendige“ virtuelle Welten erstellen, in denen Nutzer mit computergesteuerten, menschenähnlichen Charakteren, sogenannten Virtuellen Agenten (VAs), auf natürliche Weise interagieren können. Die VAs fungieren dabei meist als Wissensvermittler, beispielsweise in Form eines Guides oder Ausbilders. So können Nutzer für spezifische Aufgabenbereiche geschult werden, oder erlernen in Rehabilitationsanwendungen bestimmte motorische Fähigkeiten. Die VAs können ebenso als Schulungspartner dienen und den Nutzer bei der Ausführung schwieriger, kollaborativer Aufgaben oder bei der Verbesserung persönlicher Fähigkeiten (Soft Skills) unterstützen. Auch ist die Entwicklung und Evaluierung digitalisierter Produktionsumgebungen, sogenannter cyber-physischer Systeme, ein relevantes Beispiel: An modernen Produktionsstandorten der Plattform Industrie 4.0 oder im Krankenhaus der Zukunft wird von den Mitarbeiterinnen und Mitarbeitern erwartet, dass sie in Teams zusammenarbeiten. Zu diesen Teams gehören Menschen ebenso wie intelligente mobile Roboter. Mithilfe von VAs können solche Szenarien in einer vollständig steuerbaren virtuellen Umgebung entworfen, bewertet und trainiert werden.



Das Erstellen solcher lebensechten virtuellen Umgebungen ist jedoch eine herausfordernde Aufgabe. Überzeugende, dialogfähige VAs sind ohne die realistische visuelle und akustische Nachbildung (non-)verbalen menschlichen Verhaltens nicht denkbar. Sprache und deren Schallausbreitung im Raum, sowie Lippen-, Blick- und Körperbewegungen einschließlich Gestik und Mimik müssen in Echtzeit und auf einem hohen Qualitätsniveau simuliert werden.

Akustisch und visuell animierte Agenten

Im Profildbereich „Information & Communication Technology“ entwickeln das Institut für Technische Akustik sowie das Lehr- und Forschungsgebiet Virtuelle Realität und Immersive Visualisierung im Visual Computing Insti-



Bild 1: Zwei Forschende in der aixCAVE bei der Optimierung einer Studie zur Untersuchung der Richtcharakteristik von Sprache in einer realistischen Interaktion: Ein virtueller Agent in der Rolle eines Museumsführers verdeutlicht die Historie des visualisierten Ortes in Form eines kurzen Monologs. Die späteren Studienteilnehmenden werden dann aufgefordert, sich innerhalb der roten Fläche zu bewegen, um den Agenten aus verschiedenen Richtungen zu hören.

Foto: Peter Winandy

tute und IT Center mathematische Methoden und Algorithmen für eine audiovisuelle Simulation virtueller Welten und speziell VAs. Das Besondere an der Aachener Forschung: Wie das Wort „audiovisuell“ bereits vermuten lässt, hat neben der visuellen Animation auch die akustische Simulation eines VAs hohe Priorität. Gerade die sogenannte Auralisierung, also das „Hörbarmachen“ des gesprochenen Wortes, ist für eine realistische, ganzheitliche Simulation virtueller Agenten bedeutend. Diese gemeinsamen Arbeiten gehen in die Anfänge der 2000er Jahre zurück, als in einer Reihe gemeinsamer Projekte, die von der Deutschen Forschungsgemeinschaft gefördert wurden, die Echtzeit-Akustiksimulation sowie die 3D-Audiotechnik maßgeblich vorangetrieben wurden. Dadurch und vor allem

durch die Dokumentation und Pflege der Forschungsdaten ist die audiovisuelle VR-Simulationsplattform an der RWTH weltweit mit führend.

Ein Beispiel für die Auralisierung von VAs ist die Beachtung der Richtwirkung von Schall: Obwohl einige High-End-VR-Systeme mittlerweile über eine räumliche Audiowiedergabe verfügen, nutzen die allermeisten dieser Systeme nur eine gleichmäßige Rundum-Schallabstrahlung sowie sehr vereinfachte Raumakustikeffekte. Hierdurch können die Nutzer oder Hörer jedoch nur auf die Position eines virtuellen Sprechers im Raum schließen. Der hier entwickelte Auralisierungsansatz simuliert darüber hinaus jedoch auch die Richtwirkung der menschlichen Stimme sowie eine physikalisch basierte Akustiksimulation

der Umgebung. So kann der Hörer zusätzlich die Ausrichtung des virtuellen Sprechers akustisch erfahren, also ob Nutzer und VA von Angesicht zu Angesicht positioniert sind, oder der VA mit dem Rücken zum Hörer steht. Eine solche realistische, physikalisch basierte Akustiksimulation ist für die perfekte Illusion einer virtuellen Szene unabdingbar. Hierzu müssen zahlreiche Phänomene der Wellenausbreitung für die jeweilige Umgebung spezifisch berechnet werden. Immerhin kann jeder Mensch bereits nach wenigen Augenblicken allein aus der akustischen Information ableiten, ob man sich zum Beispiel im Freien oder in einem Raum befindet, wie groß der Raum in etwa ist, oder wie weit die Schallquelle entfernt ist.



Bild 2: Die aixCAVE am IT Center ist eine fünfseitige Virtual-Reality-Installation zur Darstellung immersiver, virtueller Umgebungen. Das Akustiksystem oberhalb der aixCAVE wird zur Beschallung des Benutzers eingesetzt und ermöglicht so eine multimodale Darbietung virtueller Umgebungen.

Foto: Peter Winandy

Um die Realitätsstreuung der akustischen Wiedergabe zu maximieren, gehen die RWTH-Wissenschaftlerinnen und Wissenschaftler noch einen Schritt weiter: Sie fügen nicht nur eine pauschale Richtwirkung hinzu, sondern berücksichtigen auch die vom VA gesprochenen Laute, Phoneme genannt. Hierzu wird in Echtzeit simuliert, wie sich die ändernde Mundform auf die gerichtete Ausbreitung des Schalls auswirkt. Hierzu spricht ein Schauspieler eine repräsentative Auswahl von Satzfragmenten. Die Bewegungen seiner Mundpartie werden dabei über Kameras und im Gesicht befestigten Markern aufgezeichnet und in einer Datenbank gespeichert. In der VR-Simulation werden diese Daten dann passend zum vom VA gesprochenen Text mit Algorithmen für eine Bewegungsvorhersage verknüpft.

Zur Erzeugung realistischer, sprachbegleitender Gesten wird ähnlich verfahren: Hierzu werden Informationen aus Datenbanken mit Techniken der inversen Kinematik kombiniert. Da dies gelegentlich zu unnatürlich aussehenden Bewegungen führt, werden auch Motion Capture Verfahren eingesetzt, wie man sie von der Charakteranimation in Filmen kennt. Dabei wird zwar die Reaktionsfähigkeit eines



Bild 3: Die Gesichtsbewegungen eines Schauspielers werden anhand mehrerer optischer Marker aufgenommen um über die sich ändernden Mundformen auch eine Phonem-basierte Richtwirkung des Schalls zu berechnen.

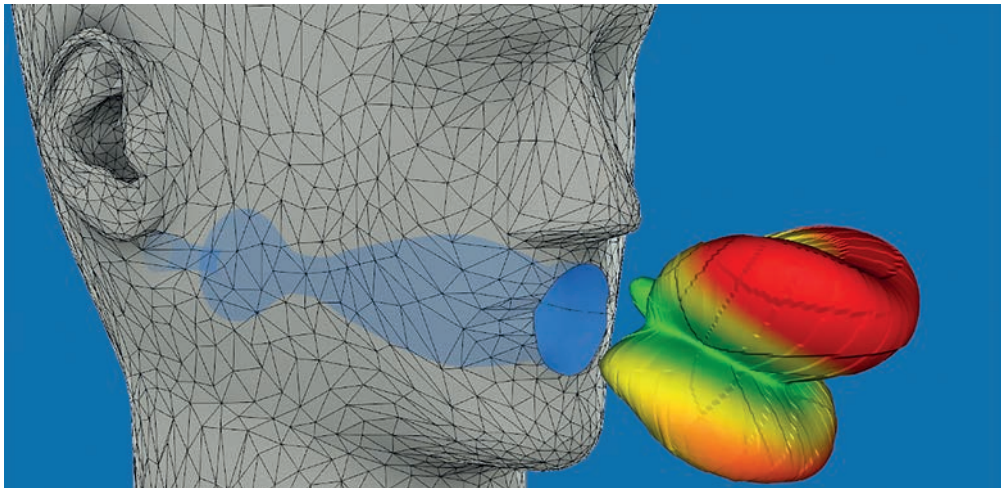


Bild 4: Eine Sprachcharakteristik (Vokal 'a' bei 1600 Hz, die Farben und die Distanz zum Ursprung beschreiben die Verstärkung/ Abschwächung in jeder Richtung) und das Kopfmodell das zur Simulation dieser genutzt wurde.

VA eingeschränkt, aber für vorbestimmte Abläufe sind die Gesten deutlich glaubhafter.

Akustische Virtuelle Realität in den Humanwissenschaften

Die Einbettung der entwickelten Methoden in realitätsnahe, audiovisuelle virtuelle Szenarien eröffnet auch Möglichkeiten für humanwissenschaftliche Disziplinen, wie der Kognitionspsychologie, der Kommunikationswissenschaft oder der Linguistik.

Verhaltensstudien bekommen in der VR einen fast natürlichen Rahmen. Trotzdem behalten die Wissenschaftlerinnen und Wissenschaftler die maximale Kontrolle über das Experiment. Insbesondere können Akteure, die in Studien zur sozialen Interaktion benötigt werden, durch VAs ersetzt werden. Dies ermöglicht ein reproduzierbares und dennoch anpassungsfähiges Verhalten der virtuellen Interaktionspartner. Audiovisuelle virtuelle Umgebungen erlauben dabei erstmals die systematische Untersuchung der Einflüsse plausibler räumlicher Akustik und visueller Reize auf die menschliche Kognition, wie die Gedächtnisleistung oder das Textverständnis.

In der Linguistik wird die audiovisuelle VR-Technologie mit VAs dazu beitragen, das Sprachverhalten in komplexen Szenen zu verstehen, die dem realen Leben nahekomen. Hierbei sind vor allem akustisch nachteilige Situationen mit mehreren Sprechern von Interesse, wie beispielsweise das Klassenzimmer oder Großraumbüros, oder auch Mehrparteienkommunikation und Szenarien im Freien. Darüber hinaus sind die VR-Umgebungen geeignet, die Interaktion von Menschen mit technologischen Systemen zu erforschen und damit die Entwicklung menschenzentrierter Technik voranzubringen.



Bild 5: Zusammenarbeit mit einem virtuellen Agenten: Ein Studienteilnehmer unterstützt den Agenten in einem virtuellen Lagerhaus bei der Suche nach bestimmten Artikeln. Hierbei werden verschiedene Richtcharakteristiken von Sprache untersucht (hier beispielhaft visualisiert).



Bild 6: Eine Benutzerin erlebt einen virtuellen Klassenraum mit zwei Agenten akustisch in einer Hörkabine.

**Deutsche Forschungsgemeinschaft
fördert Schwerpunktprogramm
AUDICTIVE**

Die RWTH-Professorinnen Janina Fels und Sabine Schlittmeier sowie -Professor Torsten Wolfgang Kuhlen haben gemeinsam mit Professor Alexander Raake von der Technischen Universität Ilmenau und Professor Steven van de Par von der Carl von Ossietzky Universität Oldenburg das Schwerpunktprogramm Auditory Cognition in Interactive Virtual Environments – AUDICTIVE – beantragt. Dieses wird

ab 2021 von der Deutschen Forschungsgemeinschaft gefördert und an der RWTH von Janina Fels koordiniert. AUDICTIVE bringt die Disziplinen der kognitiven Psychologie, der Akustik und der Informatik zusammen. Auch wird hier die interdisziplinäre Zusammenarbeit mit anderen Einrichtungen der RWTH wie dem Human Technology Center, kurz HumTec, und dem Profildbereich „Medical Science & Technology“, kurz MedST, genutzt. Ziel ist es, das Verständnis auf dem Gebiet der auditiven Kognition mit

Hilfe audiovisueller Virtueller Realität zu verbessern. Umgekehrt wird die interdisziplinäre Zusammenarbeit der Disziplinen Informatik, Akustik und Psychologie Gestaltungsrichtlinien für hochrealistische und lebendige audiovisuelle virtuelle Umgebungen in sozialen VR-Anwendungen hervorbringen. Anwendungsbereiche sind zum Beispiel Bewerbungs- und Verhandlungstrainings, das Üben öffentlicher Reden, oder die Lehrerausbildung im virtuellen Klassenzimmer. Die Projekte starten im Januar 2021.

In nicht allzu ferner Zukunft mag die Verbindung von künstlicher Intelligenz und realitäts-treuer audiovisueller VR neue Möglichkeiten eröffnen, die bis vor Kurzem nicht einmal denkbar erschienen. Mensch-Computer-Schnittstellen, in denen die Menschen mit lebensechten, intelligenten und scheinbar emotional handelnden virtuellen Agenten audiovisuell kommunizieren und interagieren, rücken in greifbare Nähe.

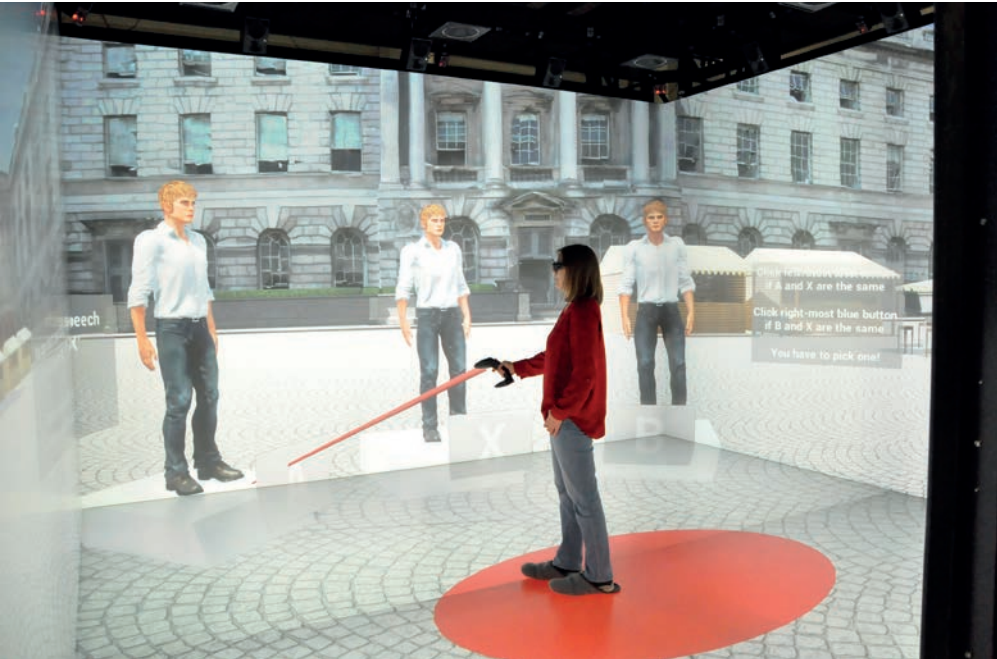


Bild 7: Beispiel eines Studienaufbaus in der aixCAVE, um in einem abstrakten Setting die Unterscheidbarkeit verschiedener Richtcharakteristiken von Sprache zu untersuchen: Zwei der drei virtuellen Agenten klingen jeweils identisch und die Aufgabe der Studienteilnehmerin ist es diese zu identifizieren.



Bild 8: Ein Schauspieler während der Aufnahme von sprachbegleitenden Gesten mithilfe eines optischen Motion Capturing Systems. Die Gesten wurden anschließend verwendet um einen virtuellen Agenten in einer Studie zu animieren.

Autoren

Andrea Bönsch, M.Sc., und Jonathan Ehret, M.Sc., sind wissenschaftliche Mitarbeiter am Lehr- und Forschungsgebiet Virtuelle Realität und Immersive Visualisierung. Univ.-Prof. Dr.-Ing. Janina Fels ist Inhaberin der Lehrstuhls für Hörtechnik und Akustik. Univ.-Prof. Dr. rer. nat. Torsten Wolfgang Kuhlen ist Leiter des Lehr- und Forschungsgebiets Virtuelle Realität und Immersive Visualisierung. Univ.-Prof. Dr. rer. nat. Michael Vorländer ist Inhaber des Lehrstuhls für Technische Akustik und Leiter des Instituts für Technische Akustik.



Bild 9: Zwei Nutzer führen eine Mehrparteienkommunikation mit einem virtuellen Agenten in einer audiovisuellen virtuellen Umgebung in der aixCAVE.
Foto: Peter Winandy

Mit „Digitalen Schatten“ Daten verdichten und darstellen

Der Exzellenzcluster „Internet der Produktion“ forscht
über die Produktionstechnik hinaus

The DFG-funded Excellence Cluster “Internet of Production” envisions a world-wide network of data-sovereign knowledge creation and data sharing in cooperation between engineering and computer science, including also business and social science aspects. This requires the integration of model-based and data-driven AI methods for coherent decision support in the development, operation, and usage cycles of production engineering. The cluster pursues a novel transdisciplinary abstraction called “Digital Shadows” to drive cross-domain research and innovation.

In der modernen Produktionstechnik sind große Mengen an Daten vorhanden. Diese sind jedoch weder einfach zugänglich noch interpretierbar oder so vernetzt, dass daraus effektiv Wissen generiert werden kann. Der RWTH-Exzellenzcluster „Internet der Produktion“, der seit Januar 2019 von der Deutschen Forschungsgemeinschaft im Rahmen der Exzellenzstrategie gefördert wird, will eine domänenübergreifende und weltweite Kooperation ermöglichen: Die Aachener informatische Infrastruktur soll die echtzeitfähige, sichere Verfügbarkeit aller relevanten Daten zu jeder Zeit an jedem Ort in semantisch angemessener und an die Diversität der Nutzenden in angepasster Qualität ermöglichen. Der Exzellenzcluster bringt so ein Ziel der

auf der Hannover Messe 2011 postulierten, aber noch lange nicht umgesetzten Vision „Industrie 4.0“ voran und ebnet den Weg in eine neue Ära der Produktion. Im RWTH-Exzellenzcluster „Integrative Produktionstechnik für Hochlohnländer“ (2006 bis 2018) wurden herausragende modellgetriebene Fortschritte in der Produktions- und Materialwissenschaft erzielt. Der hierauf aufbauende Exzellenzcluster Internet der Produktion, kurz IoP, ergänzt diese Ergebnisse jetzt nicht nur um datengetriebene Ansätze der Informatik, sondern auch um neue Managementmethoden und Geschäftsmodelle. Zur Umsetzung gehen die RWTH-Wissenschaftlerinnen und Wissenschaftler aus den Bereichen Produktionstechnik, Informatik,

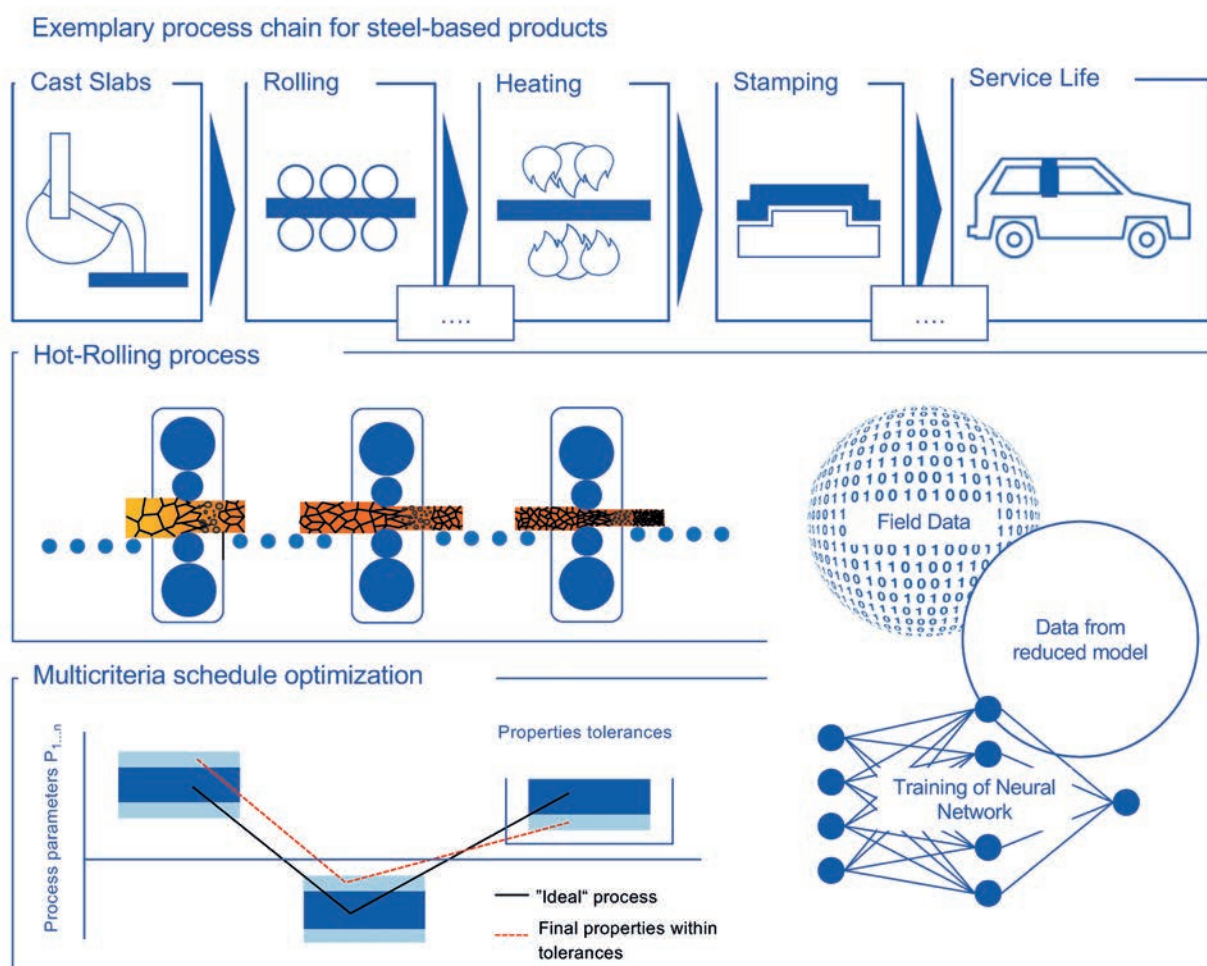


Bild 1: Einbettung mathematischer Modelle in neuronale Optimierungsnetze am Beispiel des Warmwalzens in der stahlbasierten Automobilproduktion.

Werkstoffwissenschaften sowie Wirtschafts- und Sozialwissenschaften interdisziplinäre Herausforderungen an. Die Integration von reduzierten produktionstechnischen Modellen in datengetriebenes Machine Learning dient einem domänenübergreifenden Wissensaufbau für eine kontext-angepasste und organisationsübergreifend vernetzte Produktionstechnik. Erforscht werden auch neue Methoden des ganzheitlichen Arbeitens, Stichwort New Work. Ingenieurwissenschaftliche Methoden und Prozesse werden mit sozial- und wirtschaftswissenschaftlichen Verfahren vernetzt und verbessert. Der RWTH Aachen Campus bietet mit seinen Forschungsinstituten und industriellen Partnern einzigartige infrastrukturelle Voraussetzungen

zur integrativen Entwicklung und Validierung. „Domänenübergreifende Kooperation“ bedeutet zum einen die Vernetzung von mathematischen Modellen aus unterschiedlichen ingenieur- und naturwissenschaftlichen Disziplinen. Hier sind in erster Linie Materialwissenschaft und Produktionstechnik zu verzahnen, die jeweils in einem „Cluster Research Domain“ behandelt werden. Dazu zwei Beispiele: Feinschleifmaschinen für Flugzeug- oder Kraftwerksturbinen arbeiten aufgrund modellgetriebener Steuerungsverfahren mit enormer Geschwindigkeit und Präzision, sind allerdings deshalb auch hohen Schadensrisiken bei Kohlenstoffverunreinigungen im verwendeten Stahl ausgesetzt. Es müssen daher extrem teure Spezialstähle eingesetzt

werden. Mit realzeitfähigen maschinellen Lernverfahren, welche die mathematischen Modelle um einen datengetriebenen Aspekt – die Erkennung und Umgehung von Verunreinigen im einstelligen Millisekundenbereich – ergänzen, könnten wesentlich günstigere Stähle verwendet werden.

In stahlbasierten Produktionsprozessen ist das Warmwalzen im Hinblick auf Qualitätsziele wie Materialstärke und Granularität einer der energieintensivsten Arbeitsschritte, mit einem erheblichen Anteil am CO₂-Fußabdruck der Industrie weltweit. Entscheidend für Verbesserungen ist das Scheduling der Abläufe in den großen Walzstraßen. Übliche Finite-Elemente-Simulationen brauchen für die Bewertung eines einzelnen manuell erstellten Ablaufvorschlags bis zu vier Stunden, so dass man nur sehr vorsichtig und aufgrund jahrelanger Erfahrung planen kann. Im Exzellenzcluster „Integrative Produktionstechnik für Hochlohnländer“ gelang eine Kombination reduzierter mathematischer Modelle, die bei fast gleicher Simulationsqualität die Zeit auf etwa 50 Millisekunden reduziert – ein Beschleunigungsfaktor von rund 100.000! Beschleunigt wurde allerdings nur die Bewertung eines bereits existierenden Plans, aber weder das aktive Finden guter Schedules noch die dynamische Anpassung an Kontextveränderungen – beispielsweise Veränderung der Luftfeuchtigkeit – während des Warmwalzens selbst. Im Exzellenzcluster Internet der Produktion wurde daher das mathematische Modell als Bewertungsverfahren im Training eines neuronalen Netzes („deep machine learning“) mit realen und simulierten Beispielabläufen eingebaut. Das trainierte neuronale Netz wird so zur schnellen „Suchmaschine“ für optimale Schedules. So lässt sich mit wesentlich günstigeren Toleranzen planen, weil viele Veränderungsrisiken noch zur Laufzeit abgefangen werden, und damit ein signifikanter Beitrag zur Verminderung des Klimawandels leisten.

Die Arbeiten in den einzelnen Cluster Research Domains werden über ausgewählte Anwendungsfälle verzahnt. Die Vision geht jedoch weit über die eigentlichen Produktionsprozesse hinaus: Drei weitere Cluster Research Domains behandeln daher den vorgelagerten Entwicklungsprozess, die nachgelagerte Nutzung der Produkte und das gesamtheitliche Management aller einzelnen Aspekte.

„Domänenübergreifende Kooperation“ bedeutet im Exzellenzcluster also auch kontinuierlichen Austausch zwischen den jeweils schon in sich komplizierten Netzwerken der Bereiche Entwicklung, Betrieb und Nutzung. So wird ein „World Wide Lab“ als Grundlage einer neuen Ära der Produktion angestrebt. Diese Vision trifft auf erhebliche soziale, wirtschaftliche und politische Herausforderungen. Daten müssen angepasst an die Interessen, Kulturen und Kompetenzen der enormen Diversität von Nutzern erhoben und präsentiert werden, um wirklich Wert zu schaffen. Zudem wollen viele Unternehmen möglichst viele Daten sammeln und auswerten, sind aber umgekehrt kaum bereit, Daten bereitzustellen.

Entscheidend wird es im Sinne der digitalen Souveränität aller Beteiligten sein, produktionstechnische Datenökosysteme zu entwickeln, die in hohem Maße wertschaffend sind, aber den geschaffenen Wert auch fair verteilen. Auch auf diese Forschungsfragen wird sowohl modellgetrieben als auch empirisch-datengetrieben anhand existierender allianzgetriebener Datenplattformen im industriellen Bereich eingegangen. Die Wissenschaftlerinnen und Wissenschaftler kooperieren dabei eng mit deutschen und europäischen Initiativen wie der International Data Space Association und der geplanten GAIA-X-Infrastruktur.

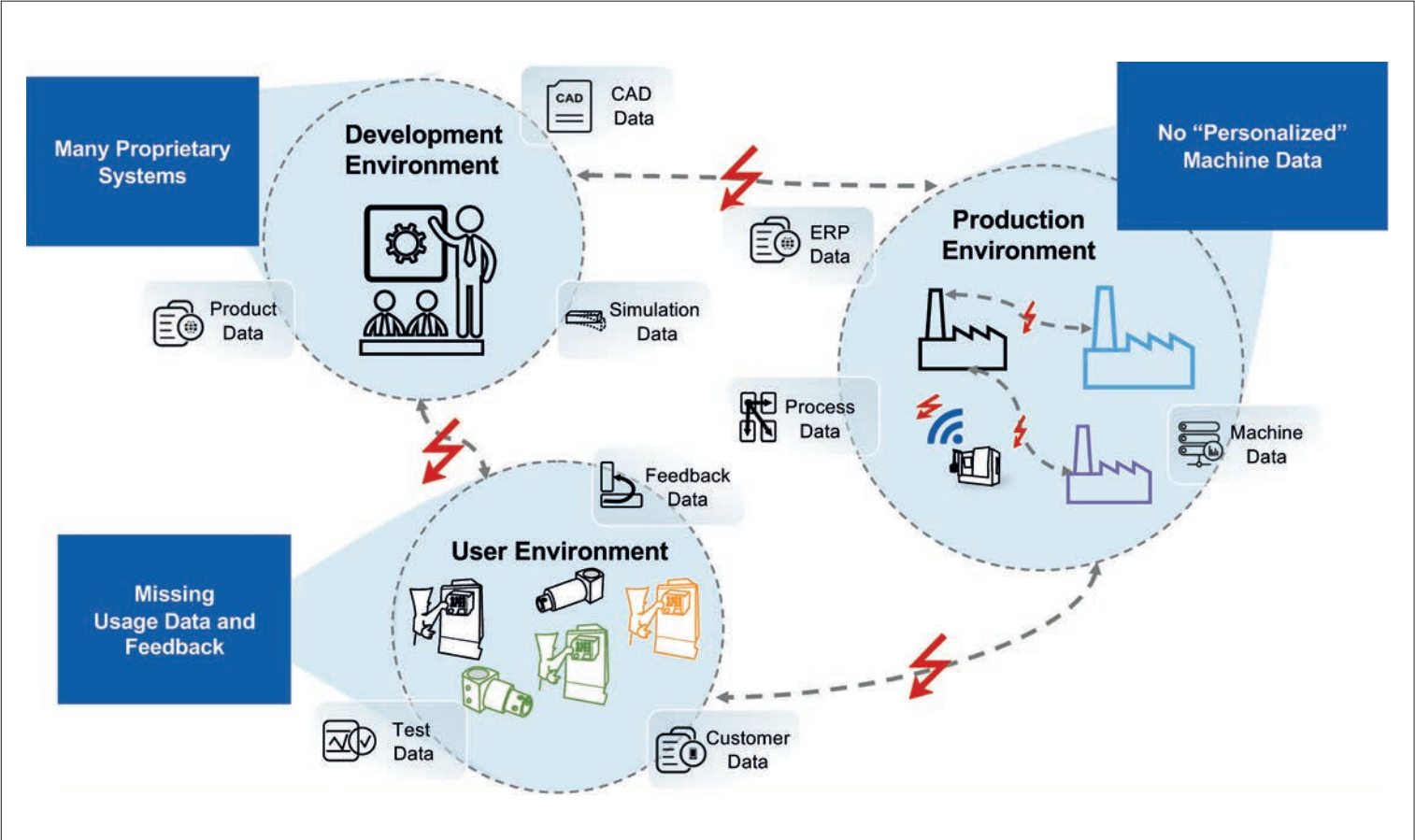


Bild 2: Der Exzellenzcluster „Internet der Produktion“ bietet Lösungen für Datenvielfalt und Integrationsprobleme beim lebenszyklusweiten Zusammenspiel zwischen Produktentwicklung, Produktionsbetrieb und Produktnutzung.

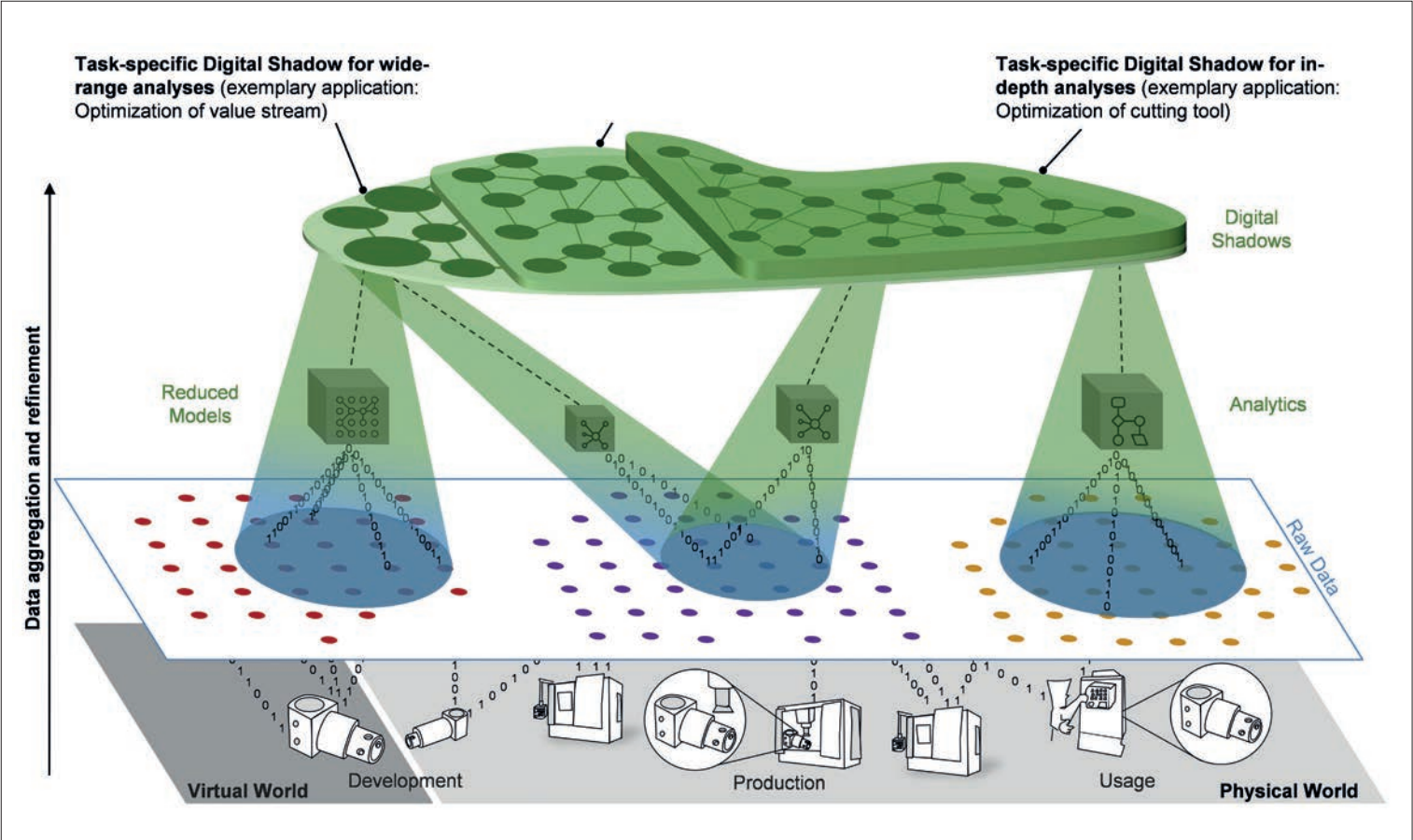


Bild 3: Digitale Schatten verdichten die riesigen Datenströme und passen damit reduzierte Modelle an die jeweiligen Aufgabenkontexte an.



Bild 4: Die Professoren Matthias Jarke (links) und Christian Brecher (rechts) koordinieren im Sprecherteam mit Professor Günther Schuh und Dr. Matthias Brockmann die interdisziplinäre Kooperation im Exzellenzcluster „Internet der Produktion“.

Foto: Peter Winandy

„Digitale Schatten“ als Kernkonzept

Viele Digitalisierungsansätze verfolgen das Konzept „digitaler Zwillinge“ – eine echtzeitfähige Simulation, die ein soziotechnisches Objekt parallel zu seiner Lebensdauer überwachend und steuernd begleitet. Der Exzellenzcluster ist allerdings zu komplex, um einen digitalen Zwilling zu realisieren. Die Infrastruktur unterstützt daher eine Vielzahl kleinerer, interagierender Objekte, „Digitale Schatten“ genannt:

- Digitale Schatten kombinieren modell- und datengetriebene interdisziplinäre Ansätze in einer jeweils kontextangepassten Weise.
- Ausgehend vom traditionellen View-Konzept im Datenbankbereich verdichten

Digitale Schatten Daten und passen deren Darstellung an die unterschiedliche Benutzerperspektiven und -fähigkeiten an.

- Die kompakte Darstellung macht es möglich, Digitale Schatten mit geringer Belastung der Kommunikationsnetze schnell zwischen maschinennahen („edge“) und zentralen Rechnern („cloud“) zu verschieben oder sogar im Netz selbst zu berechnen.
- Dadurch kann die Berechnung Digitaler Schatten an wechselnde Ressourcen und Echtzeitanforderungen angepasst werden, aber auch alle Delegationsmöglichkeiten an spezielle Hochleistungsprozessoren nutzen.

- Aus kommerzieller Sicht sind Digitale Schatten wertvolle Einheiten von Serviceaustausch und Handel, erfordern daher aber auch besondere IT-Sicherheit und Datenschutz.

Entwickelt werden Digitale Schatten für detaillierte interdisziplinäre Einzelaufgaben der Produktionstechnik und Materialwissenschaft. Diese sind aber auch in grobgranulare Management-Fragestellungen eingebettet. So unterstützt beispielsweise das „Process Mining“ auf Basis mathematischer Petrinetz-Modelle datengetriebene Analysen von riesigen Ereignismengen zur Entdeckung, Formalisierung und Compliance-Überwachung



sowohl für diskrete technische Abläufe als auch für Geschäftsprozesse. Die Dokumentation der interdisziplinären und organisationsübergreifenden semantischen Verknüpfungen erfolgt über Wissensgraphen, die auch in die Nationale Forschungsdaten-Infrastruktur für die Ingenieurwissenschaften, kurz NFDI4Ing, einfließen. Bislang lag der Schwerpunkt darauf, das Konzept des World Wide Lab am Beispiel von Fallstudien zu evaluieren und zu demonstrieren, die durch die Infrastruktur des RWTH Aachen Campus ermöglicht werden. In einem ersten Schritt wurde eine Pipeline demonstriert, die es einer örtlich verteilten Gruppe von Instituten ermöglicht, ihre unterschiedlichen Daten in

einem gemeinsamen „Data Lake“ semantisch zu verknüpfen, daraus und aus ihren jeweiligen mathematischen Methoden Digitale Schatten zu generieren und diese flexibel und sicher im Netz cloudbasiert oder maschinen-nah nutzbar zu machen. Die Wissenschaftlerinnen und Wissenschaftler arbeiten mit internationalen Partnern und führenden Infrastrukturanbietern wie Siemens Mindsphere, Amazon Web Services oder SAP Hana zusammen, um parallel zur Grundlagenforschung auch den kontinuierlichen Informationsaustausch und Technologietransfer zu sichern.

Autoren

Univ.-Prof. Dr. rer. pol. Matthias Jarke ist Inhaber des Lehrstuhls für Informatik 5 (Informationssysteme und Datenbanken) und Leiter des Fraunhofer-Instituts für Angewandte Informationstechnik FIT.

Univ.-Prof. Dr. ir. Dr. h.c. Wil van der Aalst ist Inhaber des Lehrstuhls Process and Data Science.

Dr. rer. nat. István Koren ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Informatik 5 (Informationssysteme und Datenbanken).

Univ.-Prof. Gerhard Lakemeyer, Ph.D., betreut das Lehr- und Forschungsgebiet Informatik 5 (Wissensbasierte Systeme).

Univ.-Prof. Dr. rer. nat. Bernhard Rumpe ist Inhaber des Lehrstuhls für Informatik 3 (Software Engineering).

Univ.-Prof. Dr.-Ing. Klaus Wehrle ist Inhaber des Lehrstuhls für Informatik 4 (Kommunikation und verteilte Systeme).

Univ.-Prof. Dr. phil. Martina Ziefle ist Inhaberin des Lehrstuhls für Kommunikationswissenschaft.

Univ.-Prof. Dr.-Ing. Christian Brecher ist Inhaber des Lehrstuhls für Werkzeugmaschinen und im Direktorium des Werkzeugmaschinenlabors WZL.

Univ.-Prof. Dr.-Ing. Dipl.-Wirt. Ing. Günther Schuh ist Inhaber des Lehrstuhls für Produktionssystematik und im Direktorium des Werkzeugmaschinenlabors WZL.

Dr.-Ing. Matthias Brockmann ist Geschäftsführer des Exzellenzclusters „Internet der Produktion“.



Bild 5: Interdisziplinäre Zusammenarbeit zur Implementierung des „Digitalen Schattens“ in einer produktadaptiven Roboterzelle
Foto: Peter Winandy

Materie und Licht für Quanteninformation

Forschung mit regionalen Partnern

“Matter and Light for Quantum Computing” (ML4Q) is a new Excellence Cluster funded within the Excellence Strategy by the German Research Foundation (DFG) since January 2019. It is a cooperation by the universities of Cologne, Aachen, and Bonn, as well as Forschungszentrum Jülich.

Quantum computing promises unprecedented information processing capabilities by pursuing a fundamental paradigm shift in how information is represented – using quantum states rather than discrete numbers. This pursuit requires research and technology in the fields of solid-state physics, quantum optics, and quantum information science. ML4Q brings together the best researchers from four institutions in the federal state of North Rhine-Westphalia – the universities

of Cologne, Aachen, and Bonn, as well as Forschungszentrum Jülich – to tackle the multidisciplinary issue of quantum computing. The goal of the cluster’s research is to develop new computing and networking architectures using the principles of quantum mechanics. The scientific structure of ML4Q spans four focus areas, each addressing a specific set of problems relevant to the cluster’s mission. All focus areas include theoretical as well as experimental components and transcend the boundaries of disciplines and institutions: fundamentals and technology for topological interfaces; Majorana qubits; decoherence, measurements, and error correction; and quantum connectivity. Essentially, this breakdown means that ML4Q is investigating quantum computing in terms of light, matter and information.

In Quantenrechnern wird Information durch Quantenzustände repräsentiert. Diese neue Generation von Computern könnte zur Lösung unterschiedlicher Herausforderungen beitragen, beispielsweise zur Entwicklung neuer Werkstoffe und Pharmazeutika, der Gewährleistung sicherer Kommunikation oder der Optimierung von Logistik-Arbeitsprozessen. Für solche Anwendungen würden Quantencomputer über Rechen- und Netzwerkleistungen weit jenseits existierender Technologien verfügen. Zusammen mit Forschungsgruppen der Universität zu Köln, der Rheinischen Friedrich-Wilhelms-Universität Bonn und des Forschungszentrums Jülich wollen Wissenschaftlerinnen und Wissenschaftler der RWTH die Grundlagen für den Superrechner schaffen.

Nach Jahrzehnten rasanter Entwicklung werden fundamentale Grenzen in der Miniaturisierung von Computer-Hardware erkennbar. Heutige Transistoren haben zwar eine Größe von 10nm erreicht, gehorchen jedoch noch den Gesetzen der klassischen Physik und nehmen daher klar definierbare Zustände – also 0 oder 1 – an. Bei den Mengen an Daten, die generiert werden, und den Anforderungen an Rechenleistungen gerät das Rechnen mit Bits allmählich an seine Grenzen und wird dem Bedarf an leistungsfähiger Technologie nicht mehr gerecht.

Bausteine der Quanteninformationstechnologie

Das langfristige Ziel des Clusters besteht darin, Netzwerk- und Rechenarchitekturen zu realisieren, die durch Fehlerkorrekturprotokolle geschützt und schließlich mit einer Quantenversion des Internets verbunden sind.



Bild 1: Die drei Elemente der Quanteninformationstechnologie werden im Exzellenzcluster ML4Q standortübergreifend und ganzheitlich erforscht.

Ein Blick in die Quantenwelt drängt sich auf

Unsere Welt besteht aber nicht aus Nullen und Einsen, sondern aus Quanten, die den Regeln der Quantenmechanik unterliegen – zumindest auf subatomarer Ebene. Im Quantencomputer ist der Transistor durch ein subatomares Objekt, beispielsweise ein Elektron, ersetzt. Das daraus resultierende Quantenbit, auch Qubit genannt, kann wie das Bit den Zustand 1 oder 0 annehmen – aber auch gleichzeitig im Zustand 1 und 0 sein sowie in theoretisch unendlichen Zuständen dazwischen. So können Quantenrechner Rechenschritte simultan erledigen und versprechen damit Rechenleistungen weit jenseits klassischer Computer für bestimmte Anwendungen.

Wissend um das Potenzial von Quanteninformationstechnologien als mächtige Werkzeuge beim Design neuer Materialien und Chemikalien oder der Gewährleistung sicherer Kommunikation engagieren sich Unternehmen sowie staatliche Forschungsprogramme, um Technologiekompetenz im Quantencomputing aufzubauen. Von derzeitigen Demonstratorsystemen über flexibel nutzbare Quantenprozessoren bis hin zur Vernetzung funktionsfähiger Quantencomputer im Quanteninternet wird es Jahre – vielleicht mehrere Dekaden – intensiver Forschung und die Beteiligung verschiedener Expertisen brauchen. Entwicklungen in der

Quantenmaterialforschung werden zur Herstellung robuster Qubits benötigt, die mittels der Erkenntnisse aus der Quanteninformationsforschung messbare und fehlerfreie Rechenergebnisse liefern können. Mittels der Quantenoptik können sogenannte Quantum Links realisiert werden, die die Netzwerkfunktionalität von Quantenrechnern ermöglichen. Ausgehend von diesem Paradigma ist der Exzellenzcluster „ML4Q – Matter and Light for Quantum Computing“, als Kollaboration von Wissenschaftlerinnen und Wissenschaftlern der Universität zu Köln, der Rheinischen Friedrich-Wilhelms-Universität Bonn, der RWTH sowie des Forschungszentrums Jülich entstanden. Hier wird das Ziel verfolgt, Architekturen zu schaffen, in denen fehlertolerante Quantencomputer modular realisiert und optisch miteinander vernetzt sind.

Mit welchen Qubits fängt man an?

Quanten besitzen eine Eigenschaft, die dafür verantwortlich ist, dass unsere makroskopische Welt eher klassisch als quantenhaft aussieht – sie sind äußerst störanfällig und zerfallen leicht in den klassischen Zustand. Man nennt dieses Phänomen Dekohärenz. Das Problem ist, dass aufgrund der Dekohärenz die elementaren Logikoperationen oft fehlerhaft sind – derzeit misslingt auch bei den am weitesten entwickelten Systemen typischerweise jede hundertste Operation mit zwei Qubits. Im Gegensatz dazu wird eine

elementare Rechenoperation auf unseren Desktopcomputern nur vernachlässigbar selten falsch ausgeführt. Um diese hohe Störanfälligkeit und Fehlerrate zu minimieren, werden zwei Lösungsansätze verfolgt. Im ersten versucht man bessere Qubits zu bauen; im zweiten sollen durch Fehlerkorrektur mit fehlerhaften Qubits zuverlässige Ergebnisse erzielt werden. ML4Q verfolgt in einem ganzheitlichen Ansatz beide Optionen. Dabei werden zwei Arten von Qubits ausführlich untersucht: Majorana-Qubits mit dem theoretisch vorhergesagten Potenzial, durch topologische Effekte eine bahnbrechende Immunität gegen Dekohärenzeffekte zu erzielen, und siliziumbasierte Quantenpunkte als ein mit der heutigen Halbleitertechnologie gut kompatibler Ansatz, siehe Infobox unten. Das wissenschaftliche Programm von ML4Q basiert auf vier Schwerpunktbereichen und

wird mit einem agilen Prozess komplementiert, der die kontinuierliche Aufnahme von neuen Projekten und Kollaborationen ermöglicht. Arbeitsziel des ersten Schwerpunktbereichs ist die Realisierung von Majorana-Quasiteilchen, die sich als Träger von Quanteninformation eignen, in topologischen Phasen von Festkörpern und kalten Atomen. Aus den Majorana-Quasiteilchen werden im zweiten Schwerpunktbereich Majorana-Qubits realisiert. Der dritte Schwerpunktbereich befasst sich mit Messprozessen und Konzepten zur Fehlerkorrektur, während der vierte Schwerpunktbereich die Verbindung verschiedenartiger Qubits und die Theorie von Quantennetzwerken entwickelt. Die Vision ist, möglichst optimal arbeitende Quanten-Recheneinheiten zu realisieren und die Grundlagen zu schaffen, um diese zu leistungsfähigen Systemen und Netzwerken verbinden zu können.

Quanteningenieure der Zukunft

Exzellente Forschung kann nur durch exzellente Ausbildungsprogramme den Kompetenzaufbau in der Region effektiv und nachhaltig ermöglichen. Daher ist es ein Anliegen von ML4Q, Synergien innerhalb der drei kooperierenden Universitäten zu schaffen und ein attraktives Bildungsangebot für Nachwuchswissenschaftlerinnen und -wissenschaftler zu kreieren. So bietet die RWTH die neue Vertiefungsrichtung Quantentechnologie im Masterstudiengang Physik an. Auch in der Elektrotechnik wird ein vergleichbares Studienangebot geschaffen. Zusätzlich profitiert die RWTH durch die Verbundstruktur von ML4Q von dem international etablierten Rekrutierungsprogramm der Bonn-Cologne Graduate School of Physics and Astronomy, kurz bcgs.

Neben der Kompetenzbündelung in der Region, um Nordrhein-Westfalen zu einem der Innovationsstandorte für Quantentechnologie in Europa zu entwickeln, engagiert sich ML4Q für eine bundesweite Vernetzung aller akademischen Player und sorgt damit für eine starke Stimme der universitären Grundlagenforschung. ML4Q ist seit 2019 Gründungsmitglied der Quantum Alliance. Hier haben sich fünf von der Deutschen Forschungsgemeinschaft geförderte Exzellenzcluster und weitere Zentren der Quantenwissenschaften zusammengeschlossen. Die Allianz ist zentral für die Weiterentwicklung des sich schnell entwickelnden Gebiets in Deutschland und sieht sich als einen der Motoren der Quantentechnologien sowie als Dreh- und Angelpunkt der Ausbildung in den Quantenwissenschaften.

Qubit-Realisierungen

Die Eigenschaft von Qubits, den elementaren Bauelementen von Quantencomputern, ist die Fähigkeit, bei der Speicherung und Manipulation von Zuständen möglichst ideales quantenmechanisches Verhalten zu zeigen. Insbesondere müssen sie sogenannte Superpositionszustände annehmen können, welche neben Wahrscheinlichkeiten auch den kontinuierlichen Freiheitsgrad der quantenmechanischen Phase kodieren, und miteinander verschränkt werden können. Diese Verschränkung entspricht Superpositionszuständen des Gesamtsystems, so dass zum Beispiel mehrere Lösungskandidaten einer Problemstellung gleichzeitig analysiert werden können und so eine Parallelrechnung ohne Dopplung der Hardware möglich wird.

Derzeit werden viele Ansätze zur physikalischen Implementierung von Qubits verfolgt. Trotz unterschiedlicher Entwicklungsstände ist noch nicht absehbar, welcher für eine technologische Nutzung am geeignetsten ist. Mit Lasern manipulierte Ionen und Atome im Ultrahochvakuum haben nahezu ideale Eigenschaften bewiesen, würden für die Realisierung hinreichend großer Prozessoren jedoch äußerst komplexe opto-elektronische Kontrollsysteme erfordern. In ML4Q werden diese Qubits als Modellsystem und kleine aber leistungsfähige Netzwerkelemente verfolgt. Supraleitende Qubits basieren auf wider-

standslosen elektrischen Schaltkreisen, die bei Temperaturen von wenigen Hundertstel eines Grades über dem absoluten Nullpunkt betrieben werden. Die Herstellungsverfahren sind denen der Halbleitertechnologie nicht unähnlich. Leider zeigt diese Art von Qubits eine sehr hohe Anfälligkeit gegenüber diversen Störeinflüssen. Halbleiterqubits sind der heutigen Informationstechnologie noch ähnlicher, da sie transistorartige Strukturen für das Einfangen einzelner Elektronen ebenfalls bei tiefen Temperaturen verwenden. Meist wird deren Spin-Freiheitsgrad zur Speicherung von Quanteninformation verwendet. Exzellente Eigenschaften und eine gute Robustheit einzelner Qubits konnten gezeigt werden; die Verschaltung zu größeren Einheiten ist jedoch noch eine Herausforderung, welche auch von ML4Q angegangen wird.

Majorana-Qubits beruhen auf dem theoretischen Konzept, dass sich Quantenzustände durch topologische Effekte schützen lassen. Die Grundprinzipien ähneln der Tatsache, dass sich ein Knoten in einem Seil mit festgehaltenen Enden nicht auflösen wird, auch wenn das Seil noch so sehr im Wind flattert. Jedoch ist die Demonstration bisher noch nicht geglückt und bringt große materialwissenschaftliche Herausforderungen mit sich. Diese zu meistern ist eines der zentralen Ziele von ML4Q.

Autoren

Dr. Marian Barsoum ist zuständig für die Öffentlichkeitsarbeit im Exzellenzcluster „ML4Q – Matter and Light for Quantum Computing“ und am II. Physikalischen Institut der Universität zu Köln.

Univ.-Prof. Dr. rer. nat. Hendrik Bluhm ist Inhaber des Lehrstuhls für Experimentalphysik und Standortvertreter der RWTH Aachen im Exzellenzcluster „ML4Q – Matter and Light for Quantum Computing“.

Das interdisziplinäre Graduiertenkolleg „UnRAVeL – Uncertainty and Randomness in Algorithms, Verification and Logic“

Forschung zur Eisenbahninfrastruktur und dem Eisenbahnbetrieb

Many real-world problems contain a degree of uncertainty, be it due to continuous changes, incomplete data sets, or ever-changing or unknown environments. This can have large influences on daily lives as well as hardware and software systems. Capturing and modelling the randomness and uncertainty of processes is a demanding task that requires expert knowledge from practice and science. The DFG-funded Research Training Group UnRAVeL (Uncertainty and Randomness in Algorithms, Verification and Logic) offers a stimulating and diversified environment for 12 doctoral candidates to study the effects of uncertainty and randomness. UnRAVeL

is an interdisciplinary collaboration between researchers from theoretical and applied computer science, mathematics, business administration and economics as well as railway engineering. This unique setup offers the possibility to cover problems in depth and from multiple perspectives. We present two fascinating exemplary collaborative UnRAVeL projects with a focus on railway engineering. The first project incorporates the capabilities of dynamic fault trees to estimate the criticality of infrastructure elements in railway stations. The second project deals with the theoretical complexity and efficient computation of vehicle shunting problems.

Unsicherheit wird in der Informatik immer allgegenwärtiger. Betroffen sind vor allem aktuelle Herausforderungen wie Big Data und Maschinelles Lernen, aber auch die traditionellen Fragestellungen aus den Bereichen Ereignisse und Steuerung. Anwendungen verarbeiten große Datenmengen von oft unzuverlässigen Quellen wie verrauschten Sensoren oder nicht vertrauenswürdigen Webseiten. Daten können kontinuierlichen Veränderungen unterworfen sein, in unterschiedlichen Formaten vorliegen und sind häufig unvollständig. Systeme müssen mit schwer vorhersagbaren und manchmal feindseligen Umgebungen umgehen können. Ziel des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Graduiertenkollegs „UnRAVeL – Uncertainty and Randomness in Algorithms, Verification and Logic“ ist, die probabilistische Modellierung und Analyse von Ungewissheit durch die Entwicklung neuer Theorien, Algorithmen und Verifikationstechniken signifikant voranzutreiben und auf Probleme in den Bereichen Sicherheit, Planung sowie Betriebssicherheit und Leistungsanalyse anzuwenden. Hierzu bilden Forschende

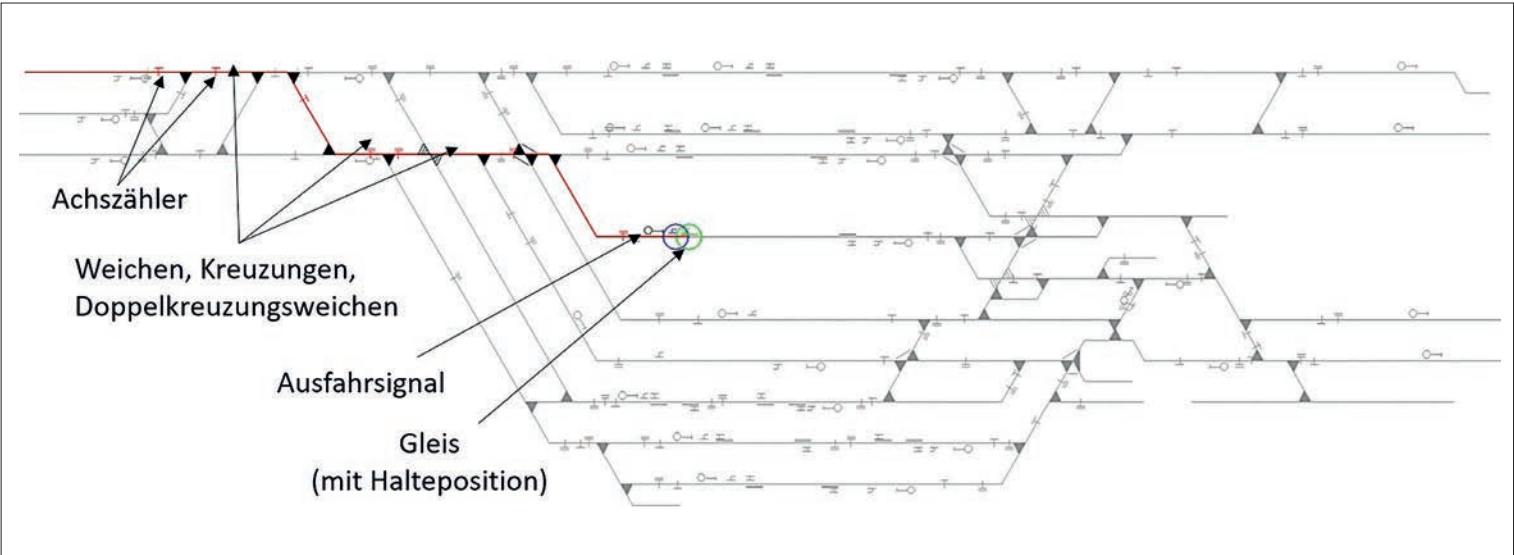


Bild 1: Darstellung der Infrastrukturelemente des Aachener Hauptbahnhofs im Spurplan



Bild 2: Die Professoren Nils Nießen (links) und Joost-Pieter Katoen (rechts) besprechen den Versuchsaufbau.
Foto: Peter Winandy

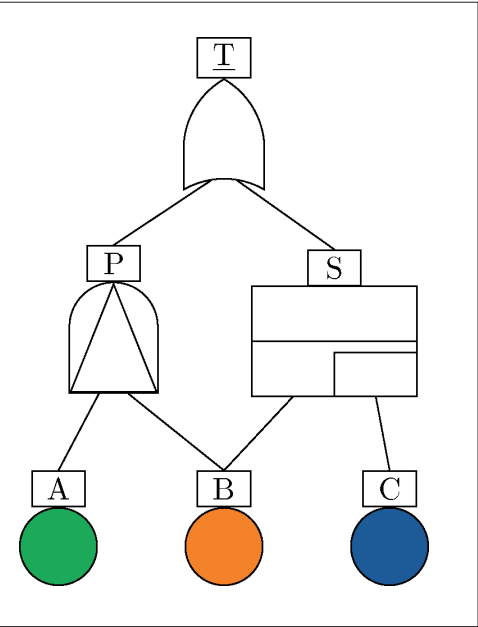


Bild 3: Ein Beispiel für einen dynamischen Fehlerbaum mit Basisereignisse (A, B, C), Priority-AND (P), SPARE (S) und OR-Gatter (T)

rinnen und Forscher der theoretischen Informatik und Mathematik aus den Gebieten computerunterstützte Verifikation, Logik und Spieltheorie, Algorithmen und Komplexität zusammen mit Expertinnen und Experten aus der Betriebswirtschaftslehre, der angewandten Informatik und dem Eisenbahnwesen den Kern des Graduiertenkollegs.

Das Graduiertenkolleg hat im Oktober 2017 den Betrieb aufgenommen und zeichnet sich durch den hohen Stellenwert von interdisziplinärer Forschung und die Fokussierung auf Unsicherheit aus.

In UnRAVeL sind zwölf Promovierende aufgenommen und durch die Deutsche Forschungsgemeinschaft finanziert. Zusätzlich gehören dem Graduiertenkolleg zwölf assoziierte Mitglieder an. Diese heterogene Gruppe erhält ein umfassendes Qualifikationsprogramm. Eine der drei Säulen ist ein individuell zugeschnittenes Study Program mit einem Katalog aus Vorlesungen, Besuchen von Summer- und Winter Schools oder regelmäßigen Treffen. Nationale und internationale Gäste bereichern den Forschungsaustausch. Der Blick über den eigenen Tellerrand wird durch Qualifikationsmaßnahmen angestoßen: Das PromotionsCAFÉ diskutiert Themen und Fragestellungen, die über den „normalen“ wissenschaftlichen Bereich hinausgehen wie Karriereplanung, Berufsaussichten, Auslandsaufenthalt, Umgang mit Minderheiten, Gender und vieles mehr. Die Grundlagen des guten wissenschaftlichen Arbeitens wie der Ablauf eines Begutachtungsprozesses, Bedeutung von Ranking- und Zitationsanalysen (h-Index),

Plagiarismus, und Forschungsdatenmanagement spielen in UnRAVeL eine wichtige Rolle. Die Exzellenz und die Zusammenarbeit des Graduiertenkollegs zeigt sich sowohl bei den betreuenden Professorinnen und Professoren, welche eine Vielzahl an Preisen und Auszeichnungen erhalten haben, als auch bei den Promovierenden. Die Mitglieder des Graduiertenkollegs haben sowohl individuell als auch in interdisziplinären Kooperationen viele Preise für die beste Veröffentlichung in internationalen Top-Konferenzen oder für die Entwicklung von Hard- und Software bei entsprechenden Wettbewerben (unter anderen Carolo-Cup) gewonnen.

Verzahnung von Theorie und Praxis am Beispiel des Eisenbahnwesens

Das Eisenbahnwesen ist ein komplexes System und stets mit den aktuellen Entwicklungen der Zeit verknüpft. Heutzutage stellen sich aufgrund der Digitalisierung viele Herausforderungen und Chancen, die auf eine Verknüpfung von Theorie und Praxis angewiesen sind. In den letzten Jahrzehnten wurde das Verständnis für Algorithmen und Daten zum Betrieb des Eisenbahnnetzes immer weiter ausgebaut. Daher bietet die Verknüpfung der Disziplinen durch das Graduiertenkolleg vielfältige Chancen, sowohl die Grundlagenforschung voranzubringen als auch eine Verbesserung des aktuellen Eisenbahnverkehrs durch Anwendung der Informatik zu erzielen.

Detektion kritischer Infrastrukturelemente in Bahnhöfen

Jeder kennt die Situation einer Verspätung bei einer Zugfahrt. Dabei stellen Infrastrukturausfälle in den Engpässen, welche sich häufig in den größeren Bahnhöfen befinden, eine der wichtigsten Ursachen für Verspätungen dar und implizieren meist starke betriebliche

Einschränkungen. Das Ausfallverhalten einzelner Infrastrukturelemente, wie Weichen, Signale oder Gleisfreimeldeeinrichtungen, ist hingegen gut verstanden und dokumentiert. Der Frage, wie diese Elemente in einem Bahnhof interagieren, sind der Lehrstuhl für Informatik 2 (Softwaremodellierung und Verifikation) sowie der Lehrstuhl für Schienenbahnwesen und Verkehrswirtschaft nachgegangen. Die Zuverlässigkeit eines Bahnhofs, siehe Bild 1, wurde dazu als Fehlerbaum modelliert, siehe Bild 3.

In einem Fehlerbaum beschreibt das sogenannte Top-Ereignis (T) den unerwünschten Zustand, über den etwas herausgefunden werden soll. In diesem Fall die eingeschränkte oder nicht vorhandene Verfügbarkeit des untersuchten Bahnhofs für den Zugverkehr. Auf der untersten Ebene befinden sich die Basisereignisse (A, B, C), welche den Ausfall der einzelnen Infrastrukturelemente symbolisieren. Zwischen Basis- und Top-Ereignis wird das Zusammenspiel der Elemente durch sogenannte Gatter (P, S und T) modelliert. Die Gatter können verschiedene Ausfallbedingungen modellieren. Damit in dem dargestellten Beispiel P einen Fehler meldet, muss zuerst das grüne und danach das orangefarbene Basisereignis eintreten. Im Eisenbahnwesen kann dies beispielsweise bedeuten, dass bei einem ersten Ereignis über einen bestimmten Abschnitt nur mit einer verminderten Geschwindigkeit gefahren werden darf und bei dem zweiten Ereignis dieser Abschnitt gesperrt werden muss.

Das Modell wurde zu Demonstrationszwecken auf die Hauptbahnhöfe Aachen und Mönchengladbach sowie den Bahnhof Herzogenrath angewendet. Dabei wurden zunächst verschiedene Ausfallraten und Reparaturzeiten für Weichen untersucht. Unter Anwendung eines Bewertungsmaßstabs

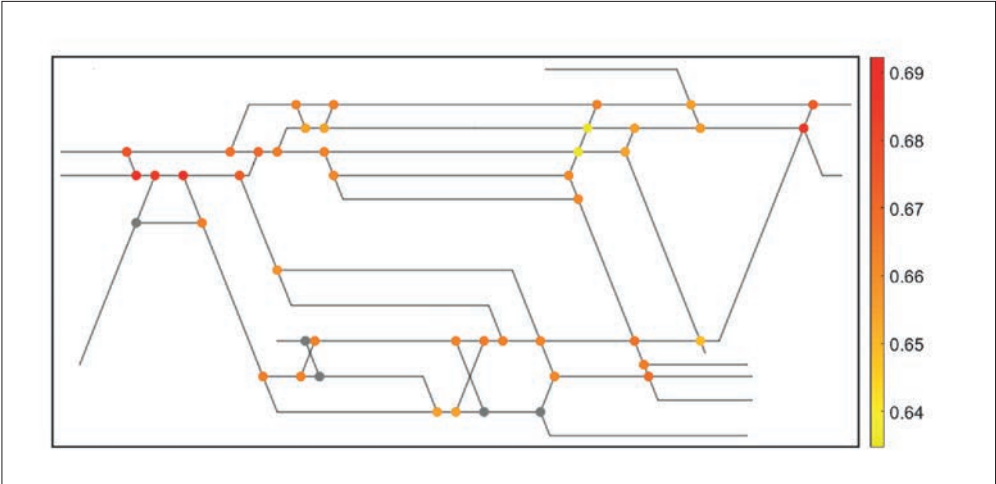


Bild 4: Die Kritikalitätsbewertung der Weichen im Hauptbahnhof Mönchengladbach

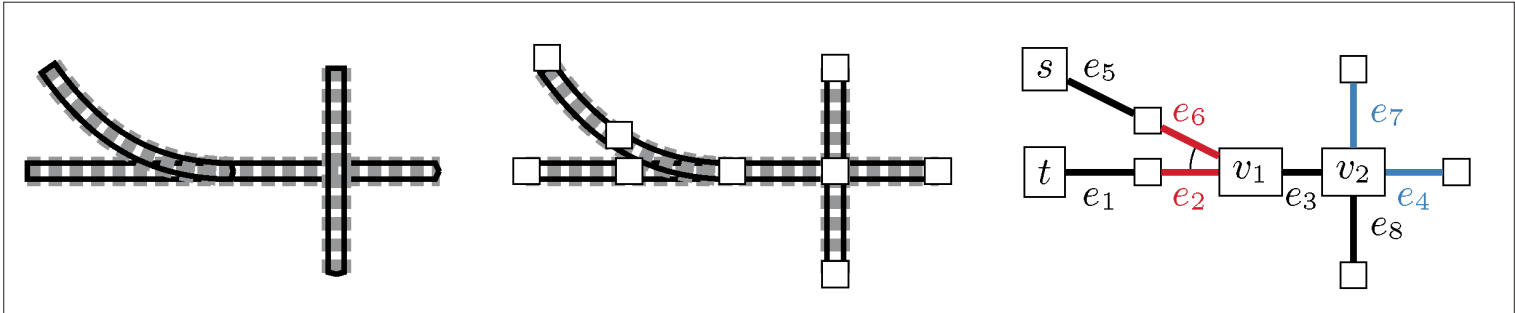


Bild 5: Die Abstraktion des Schienennetzes mittels eines Graphen

kann abschließend die Kritikalität – also die betriebliche Bedeutung – der Weichen bewertet werden. In Bild 4 werden die Kritikalitäten für ein untersuchtes Szenario für den Hauptbahnhof Mönchengladbach dargestellt. Dabei sind Weichen, die sofort zur Nichterreichbarkeit vieler Bahnsteige führen, besonders kritisch. Diese Analyse erlaubt zum einen den zielgerichteten Einsatz von Mitteln zum Infrastrukturausbau, um die Qualität im Störfall zu verbessern. Zum anderen werden so Weichen detektiert, die häufiger inspiziert und gewartet werden sollten als andere.

Erkenntnisse zur Schwierigkeit des Rangierens

Algorithmen treffen mittlerweile in vielen Bereichen unseres Lebens Entscheidungen. Sei es die Navigation im Fahrzeug, die Suche nach Begriffen im Internet oder das automatisierte Rangieren von Zügen. Mit dem effizienten Rangieren haben sich mehrere Wissenschaftlerinnen und Wissenschaftler aus UnRAVeL beschäftigt. Prinzipiell können Probleme in zwei Klassen eingeteilt werden: In solche, die einfach gelöst werden können (P) und solche, die vermutlich nicht einfach gelöst werden können (NP). Bisher haben sich alle Probleme, die NP zugeordnet werden, als praktisch schwer lösbar herausgestellt. Die Einordnung eines Problems in P oder NP hat weitreichende Konsequenzen. Während für die erste Gruppe schnelle Algorithmen gefunden werden können, die das Problem optimal lösen, sind für die zweite Gruppe entweder nur kleine Instanzen optimal lösbar oder es muss auf Näherungslösungen zurückgegriffen werden. Idealerweise kann für die Näherungslösungen eine Garantie bestimmt werden, die Aussagen wie „Der Wert unserer Lösung ist höchstens 20 Prozent schlechter als das Optimum“ ermöglichen. Um Probleme einordnen zu können, müssen reale Probleme zunächst abstrahiert und möglichst genau modelliert werden. Die Ein-

ordnung des Rangierens von Zügen haben der Lehrstuhl für Management Science, der Lehrstuhl für Informatik 1 (Algorithmen und Komplexität) sowie der Lehrstuhl für Schienenbahnwesen und Verkehrswirtschaft für verschiedene Varianten beleuchtet. Im ersten Schritt wird, wie in Bild 5, die Infrastruktur als Graph modelliert. Über diesen und die Bewegung von Einheiten auf dem Graph können dann theoretische Aussagen getroffen werden. Es zeigt sich, dass die meisten für die Realität interessanten Fälle zur Klasse NP zählen und damit nach aktuellem Kenntnisstand schwer lösbar sind. Die Beweise zur Zuordnung der Klassen beziehen sich meist auf andere Probleme, von denen bekannt ist, dass sie zu dieser Klasse gehören. Illustrativ wurde eine sogenannte Reduktion vom Brettspiel „Rush Hour“ durchgeführt. In Rush Hour soll durch geschicktes Rangieren ein gekennzeichnetes Fahrzeug aus einem Stau befreit werden, die Grundidee bei Rush Hour und dem Rangieren ist also ähnlich.

Mit dem Wissen um die Komplexität der Probleme können Algorithmen zur Unterstützung im Eisenbahnwesen zielgerichteter entwickelt werden. Die Wechselwirkung zwischen technischen Entwicklungen und der Forschung ist hochgradig relevant. Daher zeigt die interdisziplinäre Zusammenarbeit in UnRAVeL, wie positiv sich die verschiedenen Perspektiven ergänzen können. Für die Projekte wird sowohl Grundlagenforschung betrieben, zum anderen werden die gewonnenen Erkenntnisse zur Verbesserung der Anwendung, in diesem Fall des Systems Bahn, genutzt. Durch das für Investitionen stets begrenzte Budget ist ein zielgerichteter Einsatz der Mittel besonders wichtig und wird insbesondere durch die Möglichkeiten der Digitalisierung unterstützt. Hier können die Erkenntnisse aus dem Graduiertenkolleg dazu beitragen, den Wissenstransfer von der Theorie zur Anwendung zu beschleunigen.

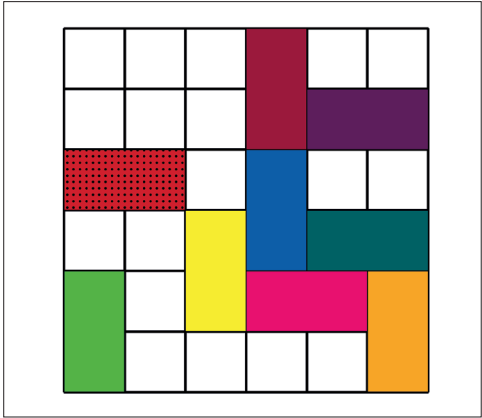


Bild 6: Das Rush Hour Puzzle

Autoren

Univ.-Prof. Dr. ir. Dr. h. c. Joost-Pieter Katoen ist Inhaber des Lehrstuhls für Informatik 2 (Softwaremodellierung und Verifikation) und Sprecher des Graduiertenkollegs UnRAVeL. Univ.-Prof. Dr.-Ing. Nils Nießen ist Inhaber des Lehrstuhls für Schienenbahnwesen und Verkehrswirtschaft und Leiter des Verkehrswissenschaftlichen Instituts. Stephan Zieger, M. Sc., ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Schienenbahnwesen und Verkehrswirtschaft.



Bild 7: Die Professoren Joost-Pieter Katoen (links) und Nils Nießen (rechts) und Mitarbeiter der Lehrstühle für Informatik 2 (Softwaremodellierung und Verifikation) beziehungsweise Schienenbahnwesen und Verkehrswirtschaft zeigen ein Modell und führen eine Testfahrt durch.

Foto: Peter Winandy



Responsible Data Science

Mit Datenwissenschaften
Entscheidungsprozesse gestalten

The availability of large amounts of data in combination with new methods of data mining and machine learning offer enormous opportunities for science, industry and society. The universal use of data science methods combined with the normative role of data-driven analyses however urgently requires an ethical perspective of the function and impact of data science in decision-making processes.

The increasing need for high-quality and robust data-driven decision-making processes in many areas of industry and society demands responsible approaches and solutions. In particular, the preservation and promotion of human autonomy needs to be carefully deliberated and designed for. RWTH Aachen University takes responsibility for the complex challenges and consequences that arise from the availability of large amounts of data and its processing for analytical purposes. While RWTH Aachen University develops and deploys data science methods in many contexts, the effects of these methods on individuals, groups and society as a whole are taken into account, evaluated and, if necessary, adjusted early and continuously. Consequently, RWTH stimulates the development of Responsible Data Science projects and initiatives in industry and society.



Große Datenmengen in Kombination mit Methoden der Datenverarbeitung und des maschinellen Lernens bieten enorme Chancen für Wissenschaft, Wirtschaft und Gesellschaft. Das interdisziplinäre Feld der Datenwissenschaften oder auch Data Science widmet sich der Entwicklung von Methoden der Datenverarbeitung und deren Anwendung. Es beschäftigt sich darüber hinaus mit der Analyse großer Datenbestände und der Gewinnung von handlungsrelevanten Erkenntnissen in einer datengetriebenen Herangehensweise. Motivation ist, wirtschaftliche und

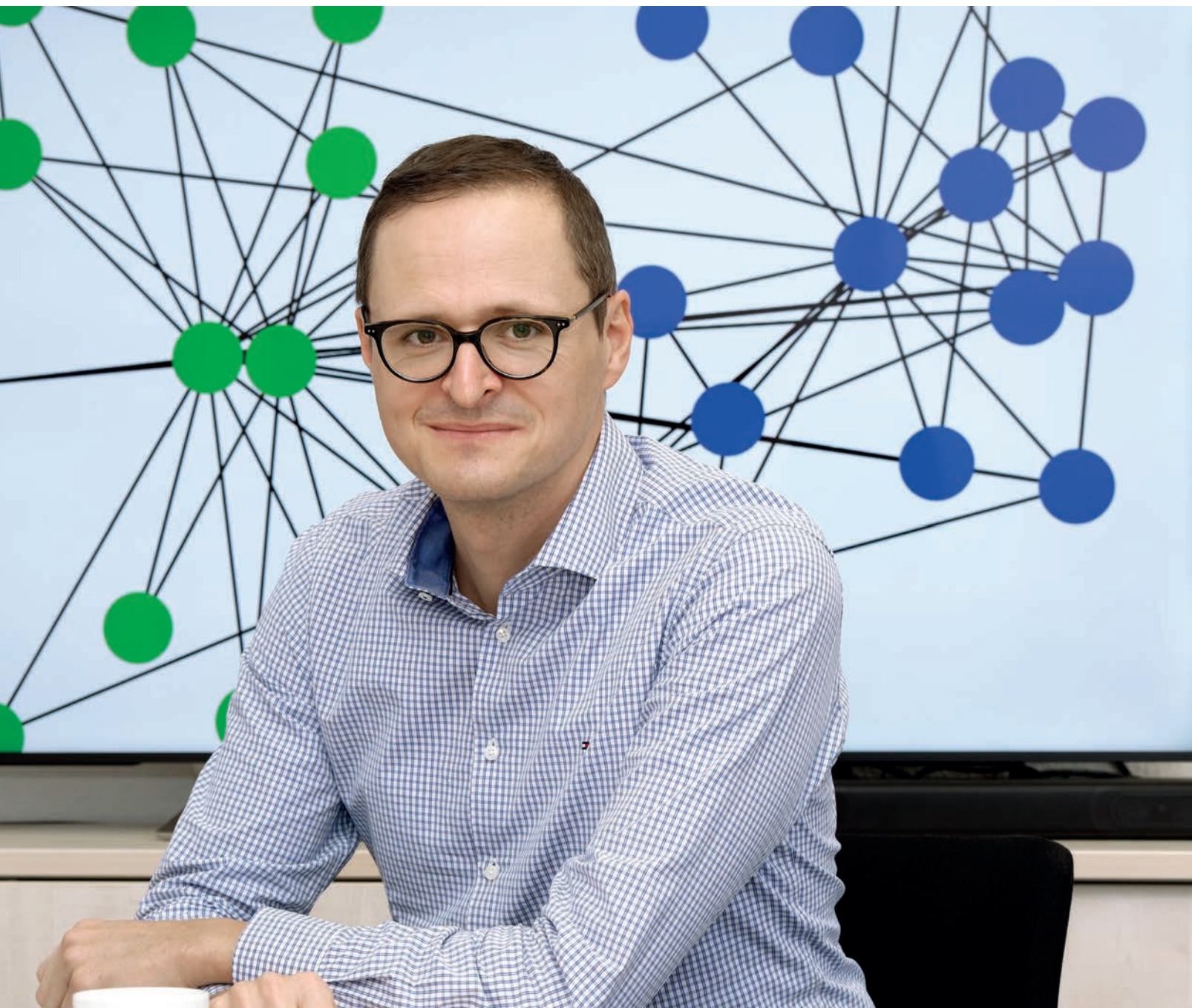


Bild 1: Professor Markus Strohmaier befasst sich unter anderem mit der verantwortlichen Implementierung und Analyse von Algorithmen für gesellschaftliche und wirtschaftliche Anwendungen.

Foto: Peter Winandy

gesellschaftliche Entscheidungen in vielfältigen Bereichen effizienter, zuverlässiger, robuster, objektiver und gerechter gestalten zu können. Die Vielzahl von Anwendungsdomänen zeugt von einem hohen Interesse am Einsatz derartiger Methoden und ihrer enormen Anziehungskraft. So werden Methoden aus Data Science beispielsweise von Einwanderungsbehörden, Justiz, Polizei, Gesundheitswesen, Kreditanbietern, Personalabteilungen oder Kinderschutzdiensten eingesetzt.

Herausforderungen für den Einsatz von Data Science

Der nahezu universelle Einsatz von Methoden aus Data Science erfordert eine frühzeitige ethische Betrachtungsweise. Eine verantwortliche Herangehensweise beginnt mit der sorgfältigen Betrachtung der Kontexte, in denen Data Science angewendet werden sollte und welche normativen Fragen durch die Anwendung entstehen. Sie berücksichtigt zudem die Frage, wann Versprechen von Objektivität und Gerechtigkeit zu rechtfertigen sind.



Bild 2: Professorin Saskia Nagel untersucht die ethischen und sozialen Konsequenzen wissenschaftlichen und technologischen Fortschritts, wie den Einfluss neuer Technologien auf unser Verständnis von Autonomie und Verantwortung.
Foto: Peter Winandy

Während beispielsweise die Verwendung neuronaler Ansätze in der Modellierung und Analyse von großen Datenmengen beeindruckende Ergebnisse in der Vorhersage oder in der Mustererkennung erzielen kann, so erhöhen diese Ansätze zugleich die Komplexität von Analyseprozessen und können zu Ergebnissen führen, die für Entscheider nicht direkt nachvollziehbar sind. Die hohe Komplexität birgt eine Reihe von Gefahren, da das Ergebnis von datenwissenschaftlich getriebenen Untersuchungen nicht immer transparent, interpretierbar, hinterfragbar oder erklärbar dargestellt wird und aufgrund stochastischer Komponenten auch nicht werden kann. Dieser Mangel wirft gesellschaftliche und ethische Fragen auf.

Ethische Fragestellungen beim Einsatz von Data Science

Wie kann nun Verantwortung im komplexen Zusammenspiel von Daten, Algorithmen und wirtschaftlichen beziehungsweise gesellschaftlichen Prozessen verstanden werden? Dies gilt insbesondere für wirtschaftlich und gesellschaftlich sensible Bereiche. Wenn zum Beispiel eine datenwissenschaftlich gestützte Anwendung in der Medizin eine fehlerhafte Diagnose stellt und ein Patient stirbt – wer ist dafür verantwortlich? Wenn ein hochautomatisiertes, weitgehend selbstfahrendes Auto einen Fußgänger anfährt, oder ein datengetriebenes Verfahren über Bildungschancen und Jobchancen entscheidet – wer trägt die Verantwortung für die Konsequenzen? Wie wird sichergestellt, dass Entscheidungen als fair wahrgenommen werden? In welchen Kontexten ist es wichtig, dass der Mensch die Prozesse im Detail verstehen kann, und welche Entscheidungen sollten Menschen vorbehalten sein? Um diesen Herausforderungen zu begegnen, ist ein Verständnis der grundlegenden normativen Fragen zu Verantwortung und Fairness unumgänglich. Datenwissenschaftlich gestützte Projekte sollten daher bereits in der Entwicklung moralisch relevante Fragen berücksichtigen. Wichtig ist vor allem die Wahrung und Förderung menschlicher Autonomie in diesen Prozessen. Dabei stellen sich Fragen zu 1) der Transparenz der Entscheidungskriterien, 2) der Erklärbarkeit der Entscheidungen, 3) der Fairness der Entscheidungen, 4) der Robustheit der Entscheidungen oder 5) der Anfechtbarkeit der Entscheidungen. Die genannten Aspekte müssen für alle – für Entwickler, Nutzer und Entscheider, für Individuen und Institutionen und für die, über die entschieden wird – nach-

vollziehbar gestaltet werden. Nur so lässt sich Vertrauen entwickeln, um die Vorteile datengetriebener Vorgehensweisen zum wirtschaftlichen und gesellschaftlichen Nutzen einsetzen zu können.

Responsible Data Science an der RWTH Aachen

Die wirtschaftlichen und gesellschaftlichen Auswirkungen datenwissenschaftlich getriebener Entscheidungen müssen regelmäßigen Evaluierungen und Regulierungen unterworfen werden. Es gilt sicherzustellen, dass die Nutzung datenwissenschaftlicher Methoden zentrale moralische Werte wie unser Bedürfnis nach Selbstbestimmung und nach Gerechtigkeit, nach Privatsphäre und nach Rechenschaft nicht beeinträchtigt oder gar untergräbt. Während die RWTH-Wissenschaftlerinnen und Wissenschaftler datenwissenschaftliche Methoden weiterentwickeln, werden die Auswirkungen dieser Methoden auf Einzelpersonen, Bevölkerungsgruppen und die Gesellschaft frühzeitig und kontinuierlich berücksichtigt, evaluiert und gegebenenfalls korrigiert. Mit zunehmenden Datenmengen wächst die Nachfrage nach technisch hochqualifizierten Datenwissenschaftlern, und mit zunehmender Nutzung von datenwissenschaftlich getriebenen Entscheidungen wächst auch die Nachfrage nach verantwortlichen und ethischen Herangehensweisen. Für eine nachhaltige Integration von Responsible Data Science in Forschung und Lehre werden gesellschaftliche und ethische Aspekte in den entsprechenden RWTH-Studiengängen berücksichtigt.

Autoren

Univ.-Prof. Dr. Saskia Nagel betreut das Lehr- und Forschungsgebiet Angewandte Ethik.

Univ.-Prof. Dr. Markus Strohmaier ist Inhaber des Lehrstuhls für Methodik und Theorie computerbasierter Geistes- und Sozialwissenschaften und wissenschaftlicher Koordinator am GESIS – Leibniz-Institut für Sozialwissenschaften.

Die Mathematik der Entscheidungen

Mit Operations Research zu bestmöglichen Handlungsempfehlungen

Our world is full of decision and planning situations of a combinatorial flavor. We seek sequences in production planning, configurations of machine setups, routes in networks. Mathematical optimization helps modeling such planning tasks and offers algorithms for solving them in a provably best possible way. When this technology is applied in business and management contexts to support decision makers, we speak of operations research. It has countless applications in production, logistics, mobility, energy, health care, education, politics, sports, and many more.

Während in der Politik noch um eine Reform des Bundestagswahlrechts gerungen wird, hat die Mathematik auf eine zentrale Frage schon eine Antwort: Wie soll das Bundesgebiet neu in Wahlkreise eingeteilt werden? Das Gesetz sieht hier nur wenige Regeln vor, etwa wie viele Wahlkreise es je Bundesland geben muss – deutschlandweit zurzeit 299 – oder dass die Wahlkreise in Bezug auf die Bevölkerungszahl etwa gleich groß sein sollen (im Mittel umfasst ein Wahlkreis fast eine Viertelmillion Deutsche). Nach Möglichkeit sollen bei der Umrandung von Wahlkreisen bereits bestehende Verwaltungsgrenzen verwendet werden. Auch muss ein Wahlkreis ein zusammenhängendes Gebiet darstellen. Fast alle Reformvorschläge sehen eine Verringerung der Anzahl der Wahlkreise vor. Dann muss Deutschland neu eingeteilt werden. Was bedeutet das? Jede Gemeinde, jeder Stadtteil muss unter Einhaltung obiger Regeln mit anderen Gemeinden zusammengefasst werden. Es ist dabei überhaupt nicht klar, wo die neuen Grenzen verlaufen werden, denn es gibt sehr, sehr viele Möglichkeiten, die weit über zehntausend Gemeinden gesetzeskonform zu gruppieren. Dies ist ein typisches kombinatorisches Problem. Zielt man gemäß dem Prinzip „eine Person, eine Stimme“ darauf ab, die Bevölkerung möglichst ausgeglichen aufzuteilen, entsteht sogar ein kombi-

natorisches Optimierungsproblem: Bestimme unter allen erlaubten Einteilungen eine, bei der die Abweichung der Bevölkerungszahl je Wahlkreis von der Durchschnittsgröße am kleinsten ist. Als zweites, laut Bundeswahlleiter gleichberechtigtes, Ziel maximiert man die Verwendung bestehender Verwaltungsgrenzen. Bild 3 zeigt, wie deutlich die Wahlkreiseinteilung 2017 in Bezug auf diese beiden Ziele verbesserbar gewesen wäre.

Die Technologie, mit der man solche Ergebnisse erzielen kann, heißt mathematische Optimierung. Sie erlaubt, den gesamten Entscheidungsspielraum zu erfassen und bezüglich einer Zielfunktion bestmögliche Lösungen zu berechnen. „Bestmöglich“ hat Bedeutungen wie am effizientesten, am schnellsten, am familienfreundlichsten. In solchen Lösungen werden sehr viele, voneinander abhängige Einzelentscheidungen simultan zu einem Gesamtkunstwerk zusammengefügt. Die Anzahl der Kombinationsmöglichkeiten wächst exponentiell, man spricht von kombinatorischer Explosion. Zwei Aspekte spielen bei deren Beherrschung eine zentrale Rolle: Entscheidungsmodelle und Optimierungsalgorithmen. Ein Entscheidungsmodell fängt ausnahmslos alle Handlungsalternativen in einer Planungsaufgabe formal präzise ein. Es besteht aus Variablen, die die einzelnen Entscheidungsbestandteile widerspiegeln und somit eine

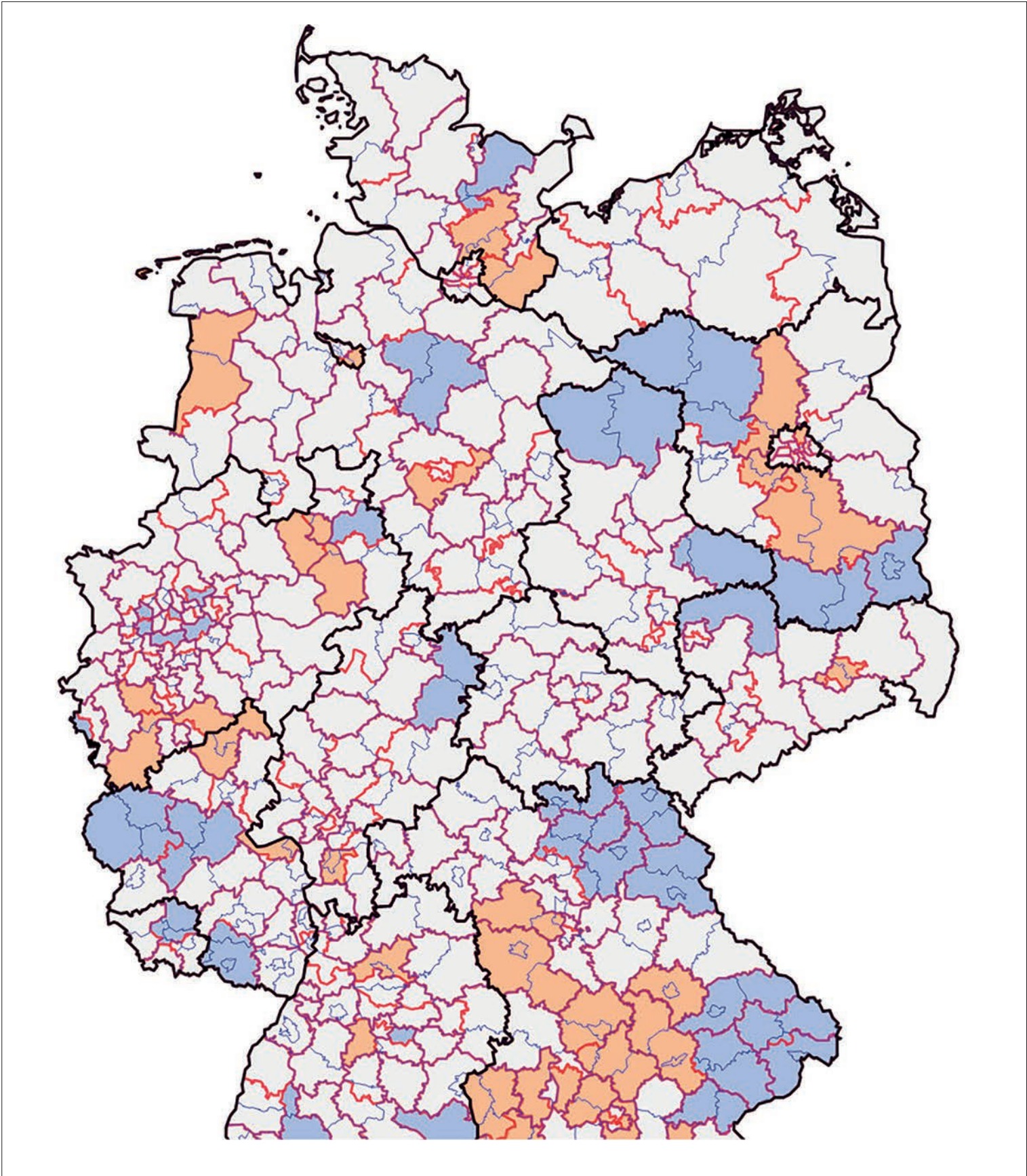


Bild 1: Einteilung des Bundesgebiets in Wahlkreise (Ausschnitt) zur Bundestagswahl 2017. Je eingefärbter ein Wahlkreis, desto größer ist die Abweichung der jeweiligen Bevölkerungszahl vom Durchschnitt. Rot sind Abweichungen nach oben, blau nach unten. In einer mathematisch optimierten Einteilung wären fast alle Wahlkreise ohne Einfärbung, die Wahlgleichheit also besser widergespiegelt.





Bild 2: Operations Research unterstützt komplexe Planungen und Entscheidungen, wie hier die bestmögliche Stapelung von Containern, mit Hilfe mathematischer Modelle und Algorithmen. Die Technologie ist so flexibel und ausgereift, dass sie in praktisch jeder Branche zum erfolgreichen Einsatz kommt.
Foto: Peter Winandy

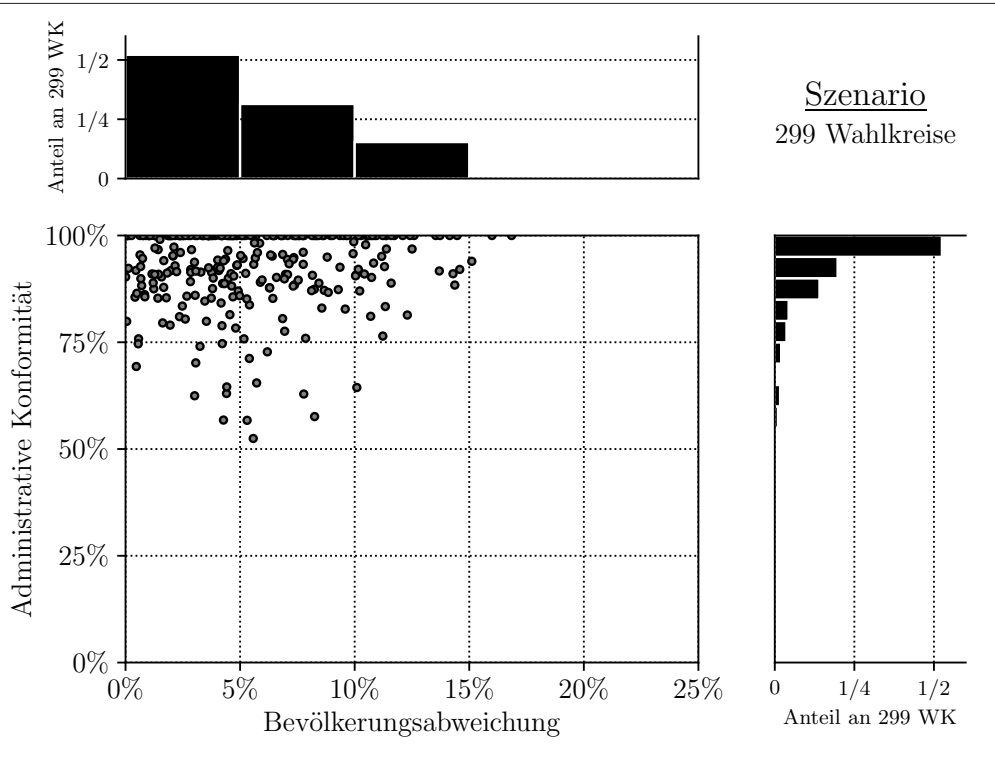
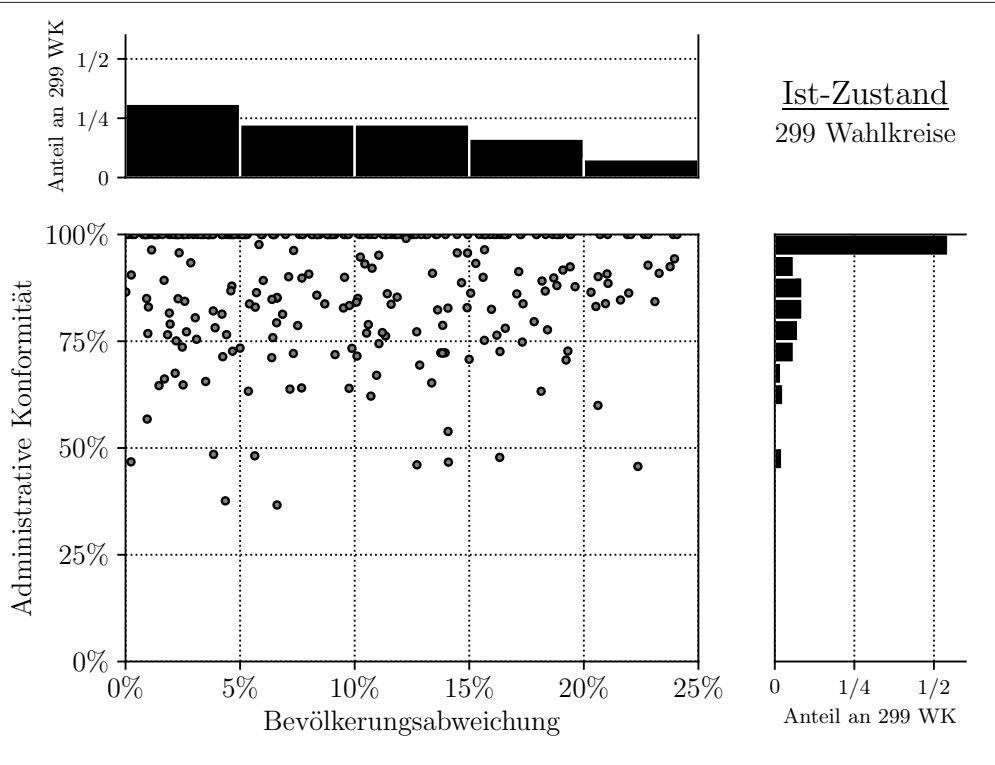


Bild 3: Auswertung der für die Bundestagswahl 2017 geltenden Einteilung in 299 Wahlkreise (oben) und einer mathematisch optimierten Einteilung (unten). Jeder Punkt entspricht einem Wahlkreis. Aufgetragen sind horizontal die Abweichungen vom Bevölkerungsdurchschnitt in Prozent (weiter links ist besser) und die Übereinstimmung der Wahlkreisgrenzen mit bestehenden Verwaltungsgrenzen in Prozent (weiter oben ist besser).

Lösung beschreiben, Restriktionen, die die Struktur einer Lösung vorgeben, und einer Zielfunktion, die die Qualität einer Lösung bewertet. Bei der Wahlkreiseinteilung gibt es je eine Variable für jede mögliche Zuordnung einer Gemeinde zu einem Wahlkreis. Diese kann nur zwei Werte annehmen, 0 oder 1, wobei 1 bedeutet, dass die Gemeinde dem Wahlkreis zugehört wird. Die Restriktionen erzwingen, dass jede Gemeinde genau einem Wahlkreis zugeordnet wird, dass die gesetzlichen Toleranzgrenzen der Bevölkerungsabweichung eingehalten werden, und dass die entstehenden Gebiete jeweils zusammenhängen. Alles dies lässt sich durch lineare Ungleichungen beschreiben. Die Zielfunktion misst zu gleichen Teilen die besagte Abweichung der Wahlkreisgrößen vom Durchschnitt und die Übereinstimmung der Wahlkreisgrenzen mit bestehenden Verwaltungsgrenzen. Es entsteht ein ganzzahliges Optimierungsmodell. Diese Modelle sind außergewöhnlich allgemein und flexibel und in beinahe allen kombinatorischen Optimierungssituationen anwendbar.

Zur Lösung der ganzzahligen Optimierungsmodelle gibt es einen sehr leistungsfähigen Algorithmus, genannt Branch-and-Cut, der für jedes derartige Modell funktioniert. Er berechnet nicht nur eine optimale Lösung, also eine bezüglich der Zielfunktionen bestmögliche Wahlkreiseinteilung, sondern liefert einen Beweis, dass es nicht besser geht. Bricht man den Algorithmus vorzeitig ab, erhält man neben einer Lösung immer auch eine Schranke, die angibt, wie weit die Lösung höchstens von einer Optimallösung entfernt ist. Diese Gütegarantien sind ein Alleinstellungsmerkmal der mathematischen Optimierung.

Der Branch-and-Cut-Algorithmus ist im Laufe von Jahrzehnten immer weiter entwickelt worden. Theoretische Erkenntnisse aus der mathematischen Forschung, auch aus Aachen, bilden hier die Grundlage. Der Algorithmus enumeriert nicht alle möglichen Lösungen durch, das würde viele Jahrhunderte dauern, sondern kann geschickt ganze Klassen von Lösungen als nicht optimal ausschließen. Dadurch ist er in der Praxis effizient. Tatsächlich haben algorithmische Verbesserungen in den letzten 25 Jahren eine Beschleunigung um einen Faktor von gut 2.000.000 bewirkt. Hinzu kommt in diesem Zeitraum die Beschleunigung handelsüblicher PCs um einen weiteren Faktor von etwa 2.000. Plakativ bedeutet dies, dass eine Optimierungsrechnung, die vor 25 Jahren noch ein Menschenleben lang gedauert hätte, heute in einer Sekun-

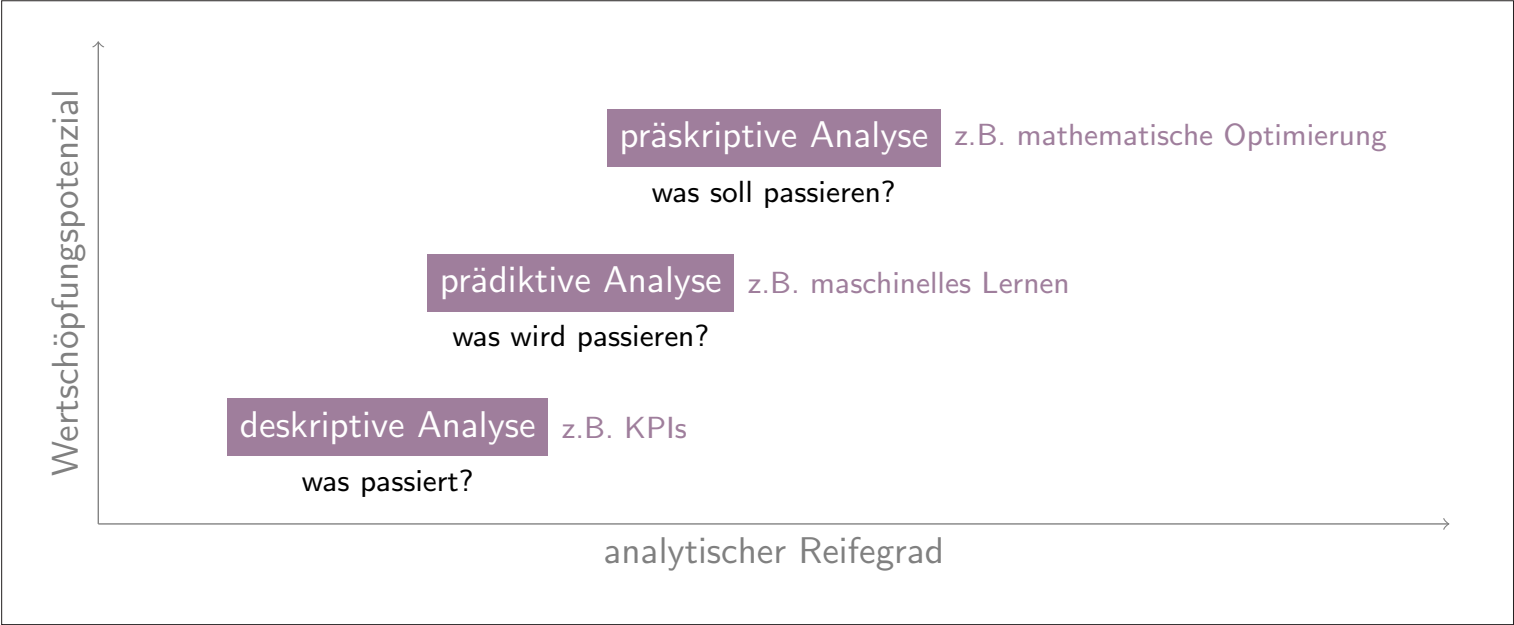


Bild 4: Arten der Datenanalyse. Mit wachsendem analytischem Reifegrad wächst in Unternehmen das Potenzial für Wertschöpfungen aus Daten. Dieser Artikel behandelt die dritte Stufe, die präskriptive Analyse.

de optimal lösbar ist. Diese Technologie in kombinatorischen Entscheidungssituationen also nicht auszuprobieren ist daher schwer begründbar. Und erweisen sich Rechenzeiten doch als inakzeptabel lang, hilft die Theorie durch Umformulierungen der Modelle und Spezialisierungen der Algorithmen. Die Vorteile einer mathematisch erzeugten Wahlkreiseinteilung liegen auf der Hand: Ausschließlich die gesetzlichen Vorgaben finden Anwendung, nur die angegebenen Ziele werden verfolgt. Wenn man von der Korrektheit der Modellierung und der Algorithmen überzeugt ist, und das dürfen wir hier unterstellen, schafft dieses Vorgehen große Transparenz. Die Flexibilität der Modelle erlaubt es, weitere Nebenbedingungen schnell einzuarbeiten. Verändert sich die Datengrundlage, und das tut sie durch demographische Entwicklung ständig, rechnet man einfach neu. Die Abstraktheit der Mathematik erweist sich als weitere Stärke. Kann man Wahlkreise optimal einteilen, so kann man das auch mit anderen Gebieten, etwa Tarifzonen, Briefzustellbezirken oder Vertriebsbereichen. In der Wertschöpfungskette von Daten kennt man mindestens drei Reifestufen der Analyse, siehe Bild 4. Die deskriptive Analyse spiegelt den Status quo wider: Was passiert gerade? Der Bereich des maschinellen Lernens mit seiner Muster- und Anomalieerkennung fällt vor allem in die prädiktive Analyse: Was passiert, wenn sich die Trends in den Daten fortsetzen? Eine konkrete Handlungsempfehlung findet erst auf der dritten Stufe statt, der präskriptiven Analyse: Was soll passieren, um

auf Basis der Daten bestmögliche Ergebnisse zu erzielen? Diese Königsdisziplin der Entscheidungen ist die Domäne des Operations Research. Am Lehrstuhl für Operations Research werden Entscheidungen in Produktion, Logistik, Mobilität, Bildung, Gesundheit, Energie, Politik und Sport unterstützt. Der internationale Stand in Forschung und Entwicklung von allgemeinen Optimierungsalgorithmen für extrem große Entscheidungsmodelle wird hier maßgeblich getrieben. Zwei Beispiele sollen die Vielfalt verdeutlichen. Der universitätsweite Stundenplan der RWTH wird jedes Semester mit Hilfe der mathematischen Optimierung geplant. Jede regelmäßige Veranstaltung benötigt einen Raum und eine wöchentliche Zeit. Die Verzahntheit interdisziplinärer Studiengänge, unterschiedliche Anforderungen der Fakultäten und Fachgruppen und eingeschränkte Verfügbarkeiten der Dozierenden sind große Herausforderungen. Das Veranstaltungsangebot ist so reichhaltig, dass Überschneidungen etwa in Wahlpflichtbereichen unvermeidbar sind, wohl aber mit mathematischer Optimierung minimiert werden können. Ganz nebenbei lassen sich so Szenarien durchspielen wie die Renovierung zentraler Hörsaalgebäude oder die Einführung neuer Studiengänge. Das Projekt „carpe diem!“ hat Vorbildcharakter für andere Universitäten. Ein aktuelles Forschungsprojekt befasst sich mit dem Lastmanagement energieintensiver Industrien. Insbesondere der wachsende Anteil erneuerbarer Energien am Strommix verursacht größere Schwankungen im An-

gebot, die auch auf der Nachfrageseite ausgeglichen werden können. So kann man die Produktion in Zeiten größeren Stromangebots verlegen und in Zeiten der Knappheit zurückfahren. Für Unternehmen stellt sich die Frage, ob sich diese Flexibilität rechnet. Tatsächlich handelt es sich auch hier um ein kombinatorisches Optimierungsproblem, bei dem die Produktion auf Basis tatsächlicher und prognostizierter Strompreise viertelstundengenau gesteuert wird. Das Projekt wird gemeinsam mit dem E.ON Energy Research Center der RWTH und dem Spin-off gapzero mathematical decision support durchgeführt.

Autor
Univ.-Prof. Dr. rer. nat. habil. Marco Lübbecke ist Inhaber des Lehrstuhls für Operations Research.

Visuelles Szenenverstehen in dynamischen Umgebungen

Deep Learning für Computer-Vision-Anwendungen

Visual scene understanding is one of the fundamental capabilities we humans possess. Perceiving the 3D world around us, recognizing objects and people that populate it, and interpreting their motion is of prime importance for our daily lives. Consequently, it is also an extremely important capability for the development of intelligent machines, such as future service robots or autonomous vehicles. If such technical systems are to move through environments built for humans, they need to be aware of their surroundings and act accordingly.

Computer Vision follows the goal of developing algorithms for the automatic analysis and understanding of the content of images and videos by computers. Historically, computer vision research has been building upon machine learning methods for a long time to achieve its goals. Still, the recent breakthrough successes in deep learning have revolutionized the field, making deep artificial neural networks the dominant tool in computer vision research. In return, computer vision has become a main testing ground for new deep learning algorithms and network architectures. This article describes current state-of-the-art research directions at the Chair of Computer Science 13 (Computer Vision).

Das Sehvermögen ist der wichtigste Sinn für uns Menschen, um unsere Umgebung wahrzunehmen, uns sicher durch sie hindurchzubewegen, und mit anderen Menschen und Gegenständen zu interagieren. Entsprechend wichtig ist ein visuelles Szenenverständnis auch für den Erfolg von technischen Geräten, wie intelligenten Fahrzeugen oder zukünftigen Servicerobotern, die ähnliche Aufgaben in komplexen Umgebungen wahrnehmen sollen.

Ziel des Lehrstuhls für Informatik 13 (Computer Vision) ist es, Algorithmen zu finden, die die beim Menschen vorhandenen Fähigkeiten zur visuellen Szenenanalyse erreichen oder übertreffen, um den Inhalt von Bildern und Videos zu interpretieren und zu verstehen. Dies ist eine sehr schwierige Aufgabe – zum einen bedingt durch die Komplexität realer Umgebungen, die eine Vielzahl an Objekten in verschiedensten Ausprägungen enthalten können; zum anderen ist für ein umfassendes Szenenverständnis das Zusammenwirken einer Vielzahl visueller Analyseschritte nötig, die wiederum unterschiedliche Verfahren benötigen.

So müssen für eine Bewegungsvorhersage anderer Verkehrsteilnehmer im Bild befindliche Objekte und semantisch bedeutsame Bildregionen erkannt und die zugehörigen Pixel vom Hintergrund segmentiert werden. Um einen Eindruck von der Bewegung eines Objekts in einer Videosequenz zu bekom-

men, müssen die Beobachtungen desselben Objekts aus den Einzelbildern miteinander assoziiert und damit die Trajektorie des Objekts durch den Raum geschätzt werden. Für eine korrekte Interpretation des Verhaltens von Personen ist zudem eine Analyse ihrer artikulierten Bewegungen sowie eine Erkennung ihrer Aktionen und sozialen Interaktionen nötig. Und all dies muss geschehen, während das Kamerasystem sich durch eine dreidimensionale und dynamisch veränderliche Umgebung bewegt und sich dadurch die Perspektive auf die betrachtete Szene ständig ändert.

Die Lösung derartiger Navigationsaufgaben durch das menschliche Gehirn ist eine gewaltige Leistung, die jedoch dadurch ermöglicht wird, dass ein signifikanter Anteil des menschlichen Gehirns der Verarbeitung visueller Informationen gewidmet ist. Entsprechend sind Aufgaben aus dem Bereich Computer Vision auch zu beliebten Test- und Anwendungsszenarien für die Entwicklung maschineller Lernverfahren geworden. Die Verwendung statistischer maschineller Lernverfahren hat in der Computer Vision eine lange Tradition und die Entwicklungen der Technik über die vergangenen beiden Jahrzehnte hinweg bedeutend geprägt. In den letzten Jahren hat insbesondere das Deep Learning immens an Bedeutung gewonnen, bei dem die zugrundeliegenden Bildmerkmale nicht mehr von menschlichen

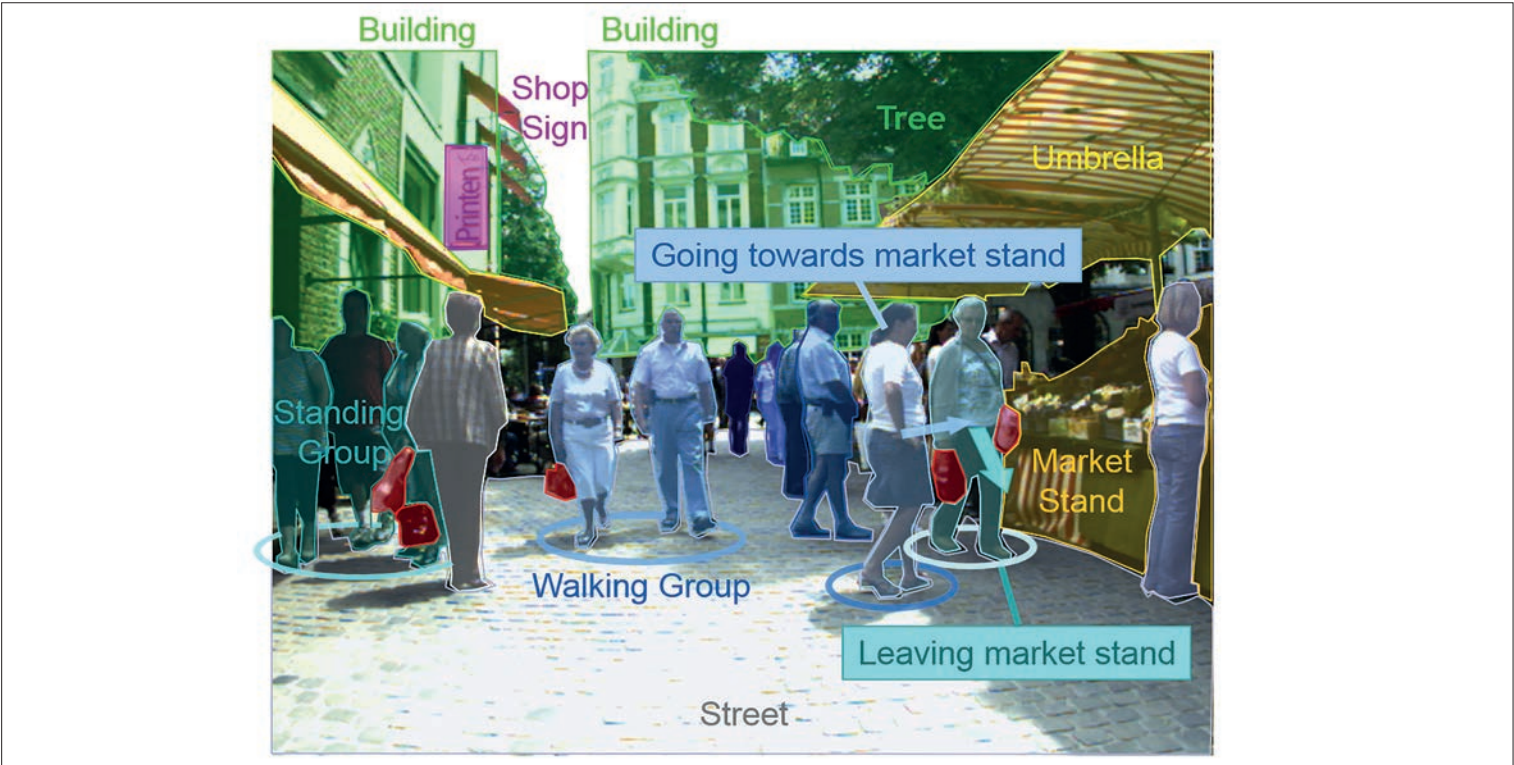


Bild 1: Die visuelle Szenenanalyse ist ein sehr komplexes Problem, für das viele Einzelaufgaben gelöst und miteinander in Bezug gesetzt werden müssen. Für eine Bewegungsvorhersage sind im Bild befindliche Objekte und semantisch bedeutsame Bildregionen zu erkennen, vom Hintergrund zu segmentieren, und über die Zeit zu tracken. Die korrekte Interpretation des Verhaltens von Personen erfordert zudem eine Analyse ihrer Körperhaltung und eine Erkennung ihrer Aktionen und Interaktionen. All diese Aufgaben werden weiter erschwert, wenn sich das Kamerasystem während der Beobachtung bewegt und sich damit die Perspektive auf die betrachtete Szene laufend ändert.



Bild 2: Vergleich der erreichbaren semantischen Segmentierungsergebnisse zweier State-of-the-art-Ansätze aus 2012 (links, [1]) und 2017 (rechts, [2], Bilder aus den jeweiligen Publikationen). Über einen Zeitraum von nur fünf Jahren ermöglichte das bessere Verständnis von Deep-Learning-Ansätzen einen Fortschritt von einer groben Klassifizierung semantischer Bildregionen hin zu einer fast pixelgenauen semantischen Interpretation komplexer Straßenszenen. (Die verschiedenen semantischen Bildregionen sind hier durch unterschiedliche Farben gekennzeichnet; die Zuordnung der Farben zu semantischen Klassen ist aber in beiden Bildern unterschiedlich).

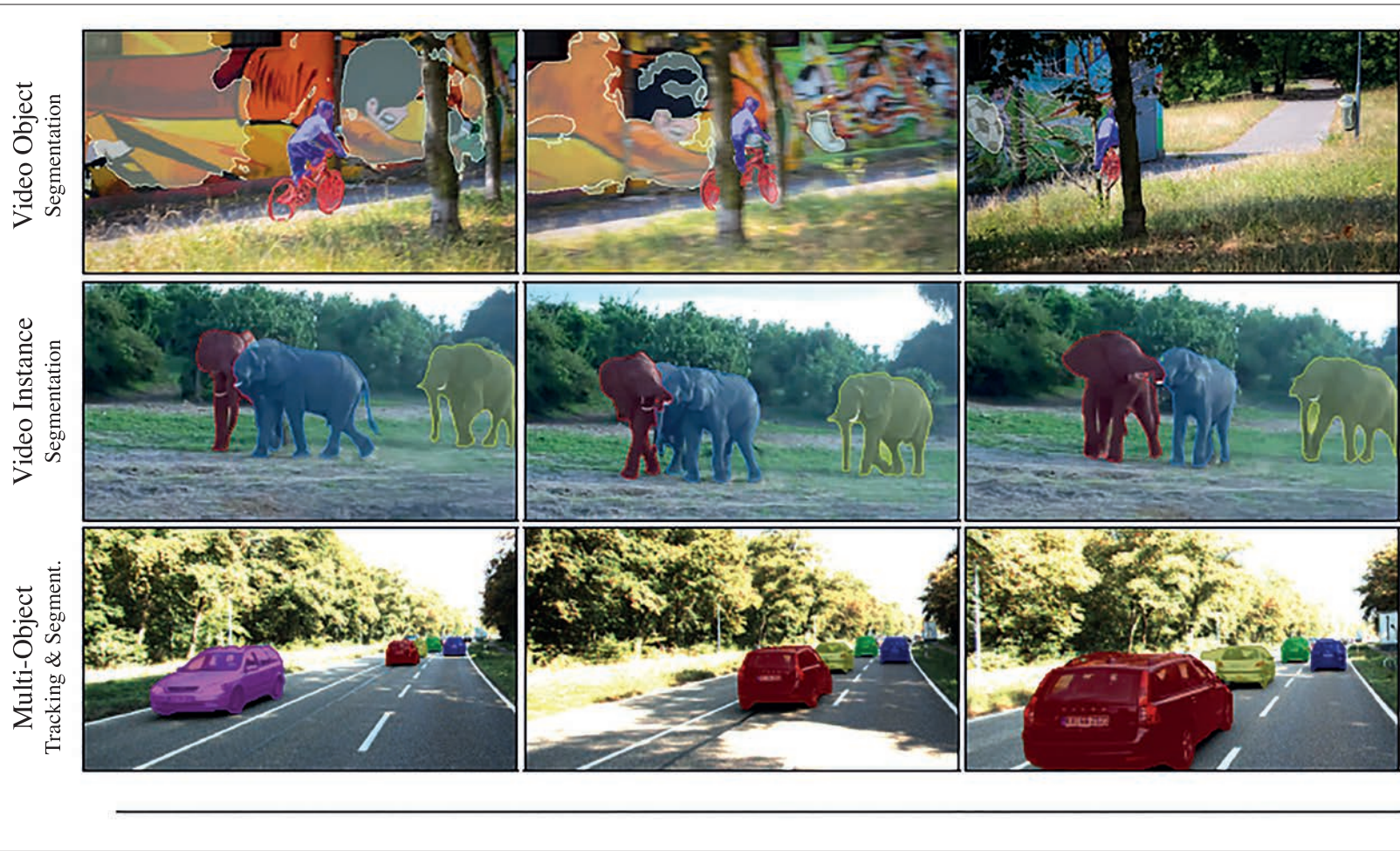


Bild 3: Die Kombination von Objekterkennung, semantischer Segmentierung und Tracking verfolgt das Ziel, bewegte Objekte in einer Videosequenz pixelgenau und zeitlich konsistent zu segmentieren. (Bilder aus [3]).

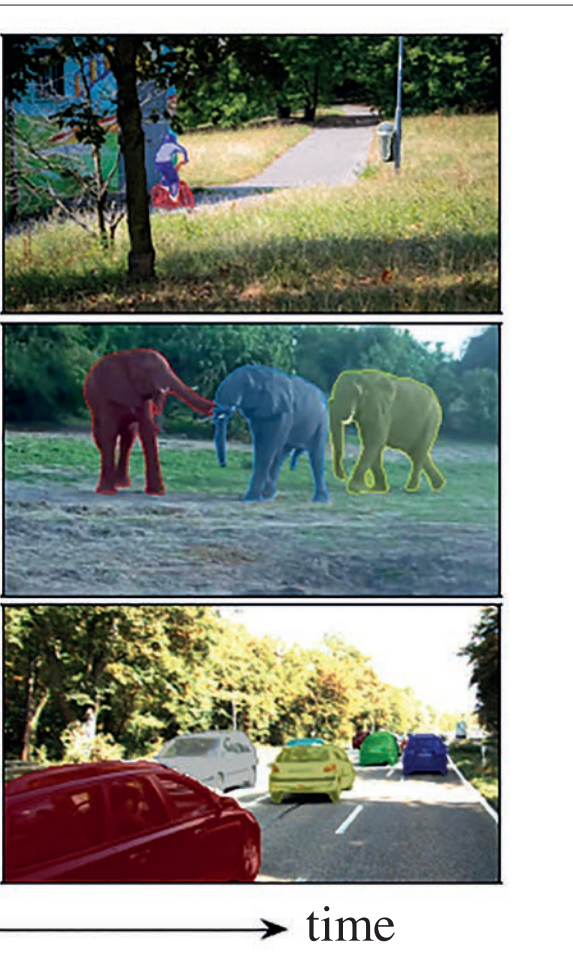
Experten entworfen werden, sondern bei dem ausgehend vom ursprünglichen Bild alle Merkmale und Verarbeitungsschritte vollständig gelernt werden. Für die Verarbeitung von Bildern sind insbesondere Convolutional Neural Networks, kurz CNNs, besonders gut geeignet. Diese besitzen gegenüber traditionellen, voll verbundenen Netzwerken, den Multi-Layer Perceptrons, den entscheidenden Vorteil, dass sie deutlich weniger gelernte Parameter benötigen, um ein tiefes Netzwerk zu definieren. Die Entwicklung von CNNs hat die Anwendung von Deep-Learning-Verfahren für Computer-Vision-Aufgaben, bei denen sehr große Bilder analysiert werden sollen, daher ermöglicht. Nichtsdestotrotz enthalten auch aktuelle CNN-Architekturen noch eine große Anzahl lernbarer Parameter, bei aktuellen Architekturen typischerweise zwischen 5 und 100 Millionen Parameter. Ein derartiges Netzwerk zu trainieren ist ein extrem hochdimensionales, nichtlineares Optimierungsproblem und erfordert sowohl ausgefeilte Optimierungsverfahren und viel Rechenaufwand als auch eine große Menge an Trainingsdaten. Die rasanten Entwicklungen im Bereich Deep

Learning und insbesondere von CNNs haben in den letzten Jahren zu immensen Fortschritten bei Computer-Vision-Verfahren geführt. Diese Fortschritte sind in Bild 2 am Beispiel der Semantischen Bildsegmentierung illustriert. Während State-of-the-art-Verfahren aus der Vor-Deep-Learning-Zeit nur eine grobe Unterteilung der Eingabebilder in semantische Bildregionen wie Auto, Fußgänger, Straße, Fassade oder Vegetation vornehmen konnten, schaffen aktuelle Verfahren eine fast pixelgenaue Interpretation hochauflöser Kamerabilder. Dies hat dazu geführt, dass die semantische Segmentierung zu einer zentralen Komponente bei der Entwicklung von Vision-Systemen für intelligente Fahrzeuge geworden ist.

Der Lehrstuhl für Informatik 13 (Computer Vision) beschäftigt sich als Teil des RWTH Centers for Artificial Intelligence mit der Erforschung und Entwicklung neuer Computer-Vision- und Machine-Learning-Algorithmen zu verschiedenen Aspekten des maschinellen Szenenverständnisses, die im Rahmen des ERC Consolidator Grants „DeeViSe – Deep

Learning for Dynamic 3D Visual Scene Understanding“ sowie in verschiedenen EU- und BMBF-Projekten verfolgt werden.

Ein Forschungsschwerpunkt ist die Kombination von Objekterkennung mit semantischer Segmentierung und Tracking in gemeinsamen, end-to-end lernbaren Ansätzen, die es ermöglichen, bewegte Objekte in einer Videosequenz pixelgenau und zeitlich konsistent zu segmentieren und vom Hintergrund zu trennen, siehe Bild 3. Derartige Verfahren haben direkte Anwendungen im Videoediting, sind aber auch als eine Vorstufe für viele weitere Computer-Vision-Aufgaben interessant. Die Herausforderung bei der Entwicklung liegt darin, effiziente Möglichkeiten zu finden, die Information über Aussehen und Form eines getrackten Objekts innerhalb der neuronalen Netzwerkarchitektur über eine gesamte Videosequenz hinweg zu propagieren, so dass dieses auch unter Bewegungen und Blickpunktänderungen noch sicher vom Hintergrund segmentiert werden kann. Ein weiterer Forschungsschwerpunkt ist die semantische Analyse von 3D-Punktwolken, wie sie bei 3D-Scans von Innenräumen, aber



auch bei der Umgebungswahrnehmung durch die LiDAR-Sensoren intelligenter Fahrzeuge anfallen. Eine Herausforderung dabei ist, dass die 3D-Punktwolken unstrukturiert und, bedingt durch die verwendeten 3D-Sensoren, meist ungleich dicht gesampelt sind (im Gegensatz zu Bildern, deren Pixel einer klaren 2D-Gitterstruktur folgen). Daher können klassische CNNs für derartige Aufgaben nicht verwendet werden. Die Arbeiten der Wissenschaftlerinnen und Wissenschaftler konzentrieren sich auf die Entwicklung von Graph Convolutional Networks mit Filteroperatoren, die sich flexibel an die 3D-Nachbarschaftsstruktur der Punktwolken anpassen, siehe Bild 4.

Ein dritter Forschungsschwerpunkt sind Algorithmen, die ein dynamisches Szenenverständnis für mobile Vision-Systeme erreichen. Sie entwickeln aus den Daten von Kameras und anderen Sensoren ein Verständnis davon, wie die 3D-Umgebung um ein intelligentes Fahrzeug oder einen Serviceroboter strukturiert ist, wie sich andere dynamische Objekte darin bewegen und leiten daraus Vorhersagen ab, wie sich diese Bewegungen

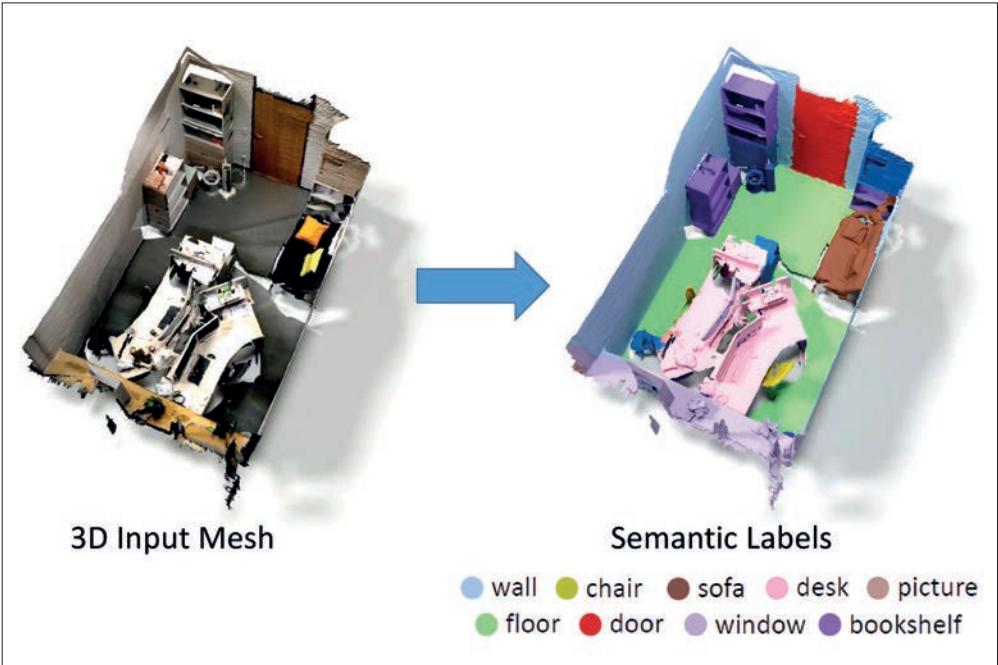


Bild 4: Ein Forschungsschwerpunkt ist die semantische Analyse von 3D-Punktwolken mittels Graph Convolutional Networks (Bilder aus [4]).

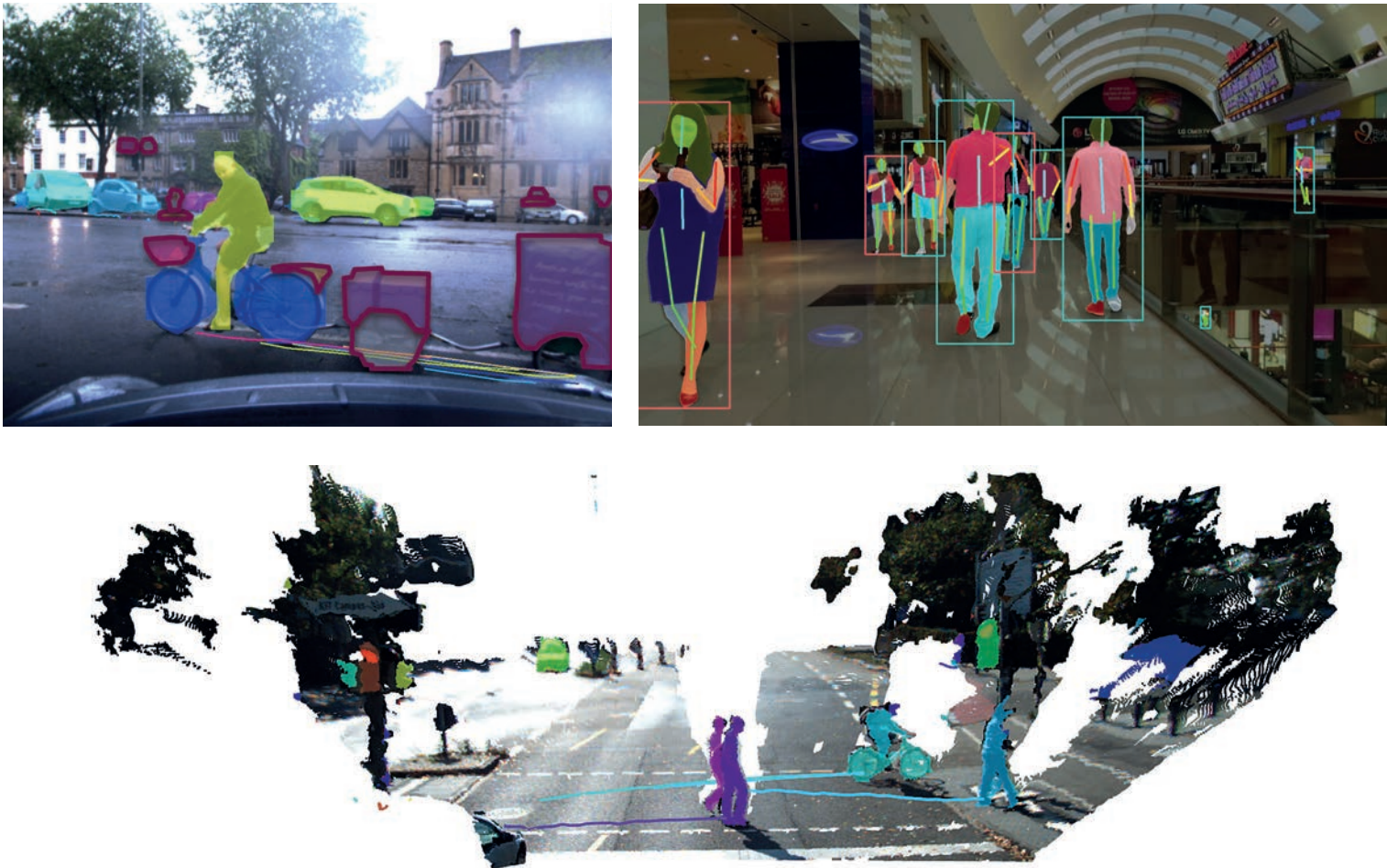


Bild 5: Beispiele für Forschungsarbeiten zu Dynamischem Szenenverstehen: (oben links) Erkennung und Tracking von „unbekannten“ Objekten im Straßenverkehr (markiert durch rote Umrandungen, Bild aus [5]); (oben rechts) Detaillierte dynamische Personenanalyse, bestehend aus Erkennung, Tracking, Segmentierung, und Körperposenschätzung (Bild aus [6]); (unten) 3D-Objektracking in Straßenszenen (die Darstellung zeigt eine aus Stereobildern rekonstruierte 3D-Punktwolke der Umgebung; getrackte Objekte sind farbig eingefärbt und mit ihrer Trajektorie dargestellt, Bild aus [5]).

in der nahen Zukunft fortsetzen werden. Ein Fokus liegt hierbei auf der Interpretation der Objektbewegungen in 3D (anstatt in Bildkoordinaten), da sich nur so die erhaltenen Ergebnisse für die Planung von Fahrmanövern nutzen lassen. Bild 5 zeigt eine Übersicht über verschiedene Aspekte des dynamischen Szenenverstehens. Zuvorderst steht das Objekttracking in 3D, um eine Interpretation der vielfältigen Bewegungen in der Szene zu bekommen, siehe Bild 5 (unten). Für die Interaktion mit Menschen ist es zudem wichtig, die menschlichen Bewegungen und Interaktionen detailliert zu analysieren. Hierzu wird an Multitask-Learning-Ansätzen geforscht, die Personenerkennung mit Segmentierung und Körperposenschätzung verbinden, siehe Bild 5 (oben rechts). In Straßenszenen ist es darüber hinaus wichtig, dass die Bildanalyse sich nicht nur auf bekannte Objekte erstreckt, sondern dass insbesondere auch unbekannte dynamische Objekte als Hindernisse erkannt und getrackt

werden, siehe Bild 5 (oben links). Hierzu wurde ein Ansatz entwickelt, der sowohl bekannte als auch unbekannte Objekte in solchen Szenarien robust tracken kann. Neben der Erforschung grundlegender Computer-Vision- und Machine-Learning-Algorithmen steht auch die praktische Anwendung dieser Algorithmen auf mobilen Robotern im Vordergrund. Die Herausforderung hierbei ist, die entwickelten Algorithmen so weit herunterzukalibrieren, dass sie auch auf der stark limitierten Hardware eines mobilen Roboters umgesetzt werden können und dabei die hohen Anforderungen an die Laufzeit einer Online-Sensordatenauswertung erfüllen. Hierfür wurde ein mobiles Vision-System zum Personentracking mittels Kamera- und LiDAR-Sensoren entwickelt, das bereits in den EU-Projekten SPENCER und STRANDS sowie dem BMBF-Projekt FRAME erfolgreich auf verschiedenen Roboterplattformen eingesetzt wurde. Unter anderem mit Hilfe dieses Vision-Systems bestand der SPENCER-

Roboter bereits 2016 erfolgreich seinen Praxistest, in Amsterdams Flughafen Schiphol Passagiere durch ein dicht frequentiertes Terminal zu ihren Abfluggates zu führen und dabei anderen Fußgängern auszuweichen, siehe Bild 6. Im aktuellen EU-Projekt „Safe Navigation of Robots in Dense Human Crowds“, kurz CROWDBOT, wird das Vision-System verbessert und erweitert. Das Projekt verfolgt das Ziel, Technologien für mobile Roboter zu entwickeln, die sich durch dichte Menschenmengen bewegen können, bei denen ein physischer Kontakt zwischen Roboter und Mensch nicht mehr vermieden werden kann. Diese Fähigkeiten sind beispielsweise für Evakuierungsszenarien wichtig, in denen ein autonom fahrender Rollstuhl sich auch in dichtem Gedränge weiterbewegen muss und den Verkehrsfluss nicht blockieren darf.

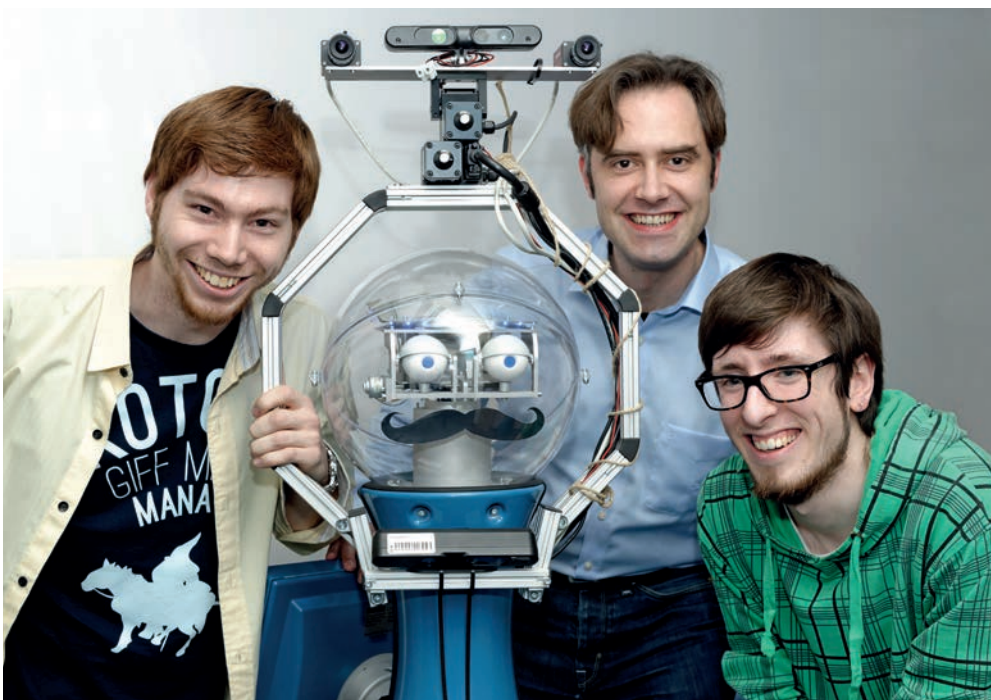


Bild 6: Das Vision System zum Personentracking in dynamischen Umgebungen wurde bereits in den EU-Projekten SPENCER (oben, Foto: © 2016 KLM [7]) und STRANDS (unten, Foto: Peter Winandy) auf verschiedenen Roboterplattformen erfolgreich eingesetzt.

Literatur

- [1] Floros, G., Leibe, B., Joint 2D-3D Temporally Consistent Segmentation of Street Scenes, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012.
- [2] Pohlen, T., Hermans, A., Mathias, M., Leibe, B., Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017.
- [3] Athar, A., Mahadevan, S., Osep, A., Leal-Taixe, L., Leibe, B., STEm-Seg: Spatio-Temporal Embeddings for Instance Segmentation in Videos, European Conference on Computer Vision (ECCV), Springer, 2020.
- [4] Schult, J., Engelmann, F., Kontogianni, T., Leibe, B., DualConvMesh-Net: Joint Geodesic and Euclidean Convolutions on 3D Meshes, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020.
- [5] Osep, A., Voigtlaender, P., Weber, M., Luiten, J., Leibe, B., 4D Generic Video Object Proposals, IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2019.
- [6] Pfeiffer, K., Hermans, A., Sarandi, I., Weber, M., Leibe, B., Visual Person Understanding through Multi-Task and Multi-Dataset Learning, German Conference on Pattern Recognition (GCPR), Springer, 2019.
- [7] EU Project SPENCER Press Material, <http://www.spencer.eu/press.html>

Autor

Univ.-Prof. Dr. sc. tech. Bastian Leibe ist Inhaber des Lehrstuhls für Informatik 13 (Computer Vision).

Sprachverarbeitung und künstliche Intelligenz

Speech and language are typically considered to be inherently human capabilities. With the advent of computing machinery, the automatic processing of speech and language became one of the cornerstone goals in artificial intelligence. Typical tasks involve the recognition and comprehension of spoken language (speech signal) and translation between languages (written text). The unifying term for this area is human language technology.

The most successful approaches to speech and language processing to date are based on the idea that a computer learns from (possibly a huge amount of) example data and uses plausibility assessments rather than externally provided categorical rules. Such approaches are based on statistical decision theory and machine learning. During the last two decades, artificial neural networks along with deep learning were found to provide a very powerful tool for machine learning and human language technology.

Die Verarbeitung gesprochener und geschriebener Sprache gilt als typisch menschliche Fähigkeit. Mit dem Aufkommen der Computer wurde es als eine der Kernaufgaben der sogenannten künstlichen Intelligenz angesehen, automatische Systeme für die Sprachverarbeitung zu entwickeln. Typische Aufgaben der Sprachverarbeitung sind das Erkennen und das Verstehen gesprochener Sprache und die Übersetzung zwischen Sprachen. Allgemein bezeichnet der Begriff Sprachverarbeitung (human language technology) jede

Art der Verarbeitung von gesprochener Sprache (Sprachsignal, speech) und geschriebener Sprache (digitaler Text, language). Keine Innovation hat das menschliche Leben mehr verändert als die Entwicklung der Sprache und die Erfindung der Schrift. Eine entscheidende Fortführung dieser Entwicklungsrichtung war der Buchdruck, und in Zukunft wird dasselbe auf die automatische Verarbeitung von geschriebener und gesprochener Sprache zutreffen. Durch die Entwicklung in der Informationstechnologie, insbesondere

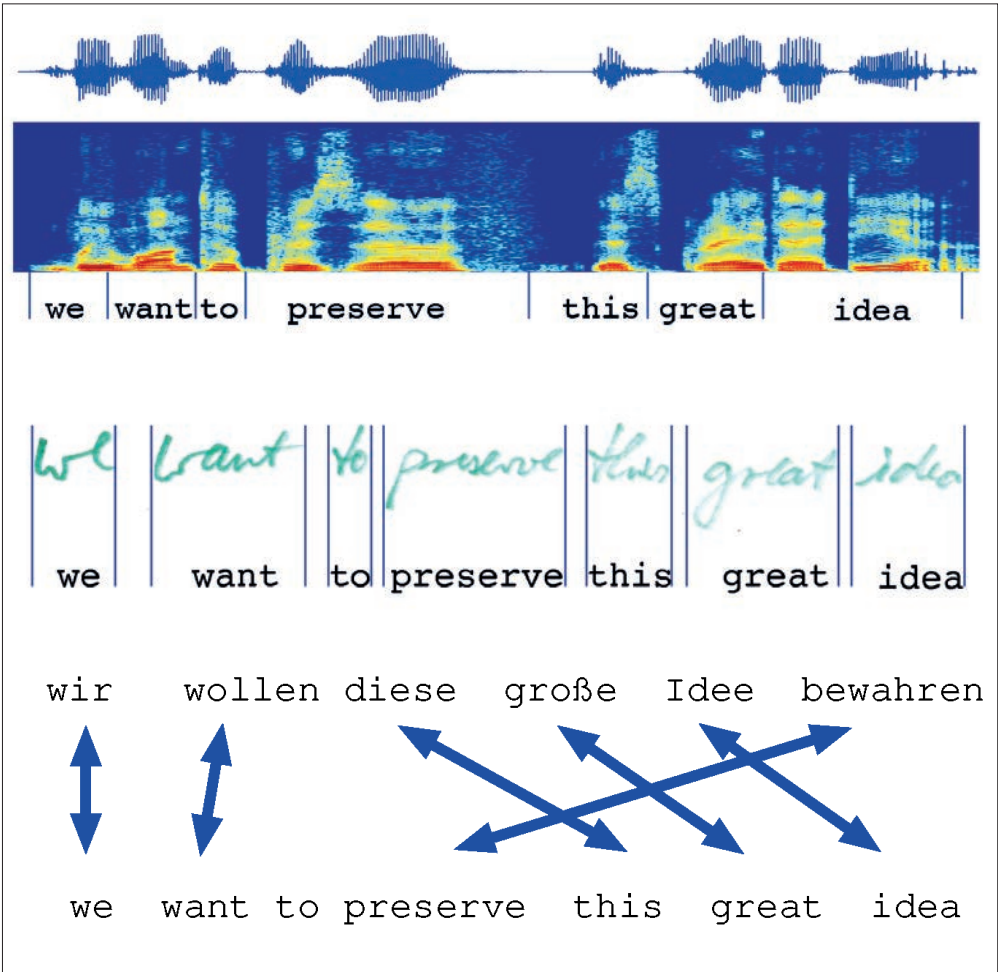


Bild 1: Drei Aufgaben der maschinellen Sprachverarbeitung: Spracherkennung, Handschrifterkennung, Übersetzung.

durch die modernen Speichermedien und die weltweite Vernetzung der Rechner im Internet inklusive des World Wide Web, werden immer größere Mengen von Text- und Multimedia-Dokumenten von jedem Handy, Laptop oder PC aus zugänglich. Neben reinem Text enthalten diese Dokumente Bilder, Audiosignale und Videoaufzeichnungen. Zu der Verarbeitung der geschriebenen Sprache (Texte) und der gesprochenen Sprache (Audiosignale) kommt jetzt die Verarbeitung von Bildinformationen hinzu. Außerdem kann es sich um Dokumente in unterschiedlichen Sprachen handeln; Multilingualität und die Übersetzung zwischen Sprachen sind also weitere wichtige Aspekte. Der potenzielle Nutzer steht vor dem Problem, aus der Fülle des Materials zielgerichtet das für ihn Wichtige auszuwählen. Eine der Herausforderungen an die Sprachverarbeitung und das maschinelle Lernen ist die Entwicklung von automatischen Verfahren, die diese Informationsflut bewältigen und dem Benutzer einen einfachen Zugang zu den gewünschten Informationen ermöglichen. Neben diesem Informationszugang ist für den Menschen die wichtigste Funktion der Sprache die direkte Kommunikation. Die Bedienung vieler Systeme wie Handy und PC, Geräte und Maschinen würde erheblich einfacher, und viele interessante Anwendungen würden sich ergeben, wenn der Computer die gesprochene Sprache (fast) perfekt erkennen und verstehen könnte. Genauso könnte die Sprachbarriere zwischen Menschen mit unterschiedlichen Muttersprachen durch eine automatische Übersetzung gesprochener Sprache beseitigt werden. Aufgrund dieser vielen Anwendungen wird an der maschinellen Sprachverarbeitung in Forschungseinrichtungen weltweit gearbeitet. Hier sind vor allem die Labors der großen IT-Firmen zu nennen und darunter insbesondere die GAFAMs (GAFAM: Google, Amazon,

Facebook, Apple, Microsoft). Weltweit gibt es einen ständig wachsenden Bedarf an Expertinnen und Experten in diesem Bereich. In Aachen gibt es mittlerweile mehrere IT-Firmen (Amazon, Apple, AppTek, Ebay, Nuance/Cerence), die die Nähe zur RWTH und zu ihren Absolventinnen sowie Absolventen gerne nutzen.

Warum ist maschinelle Sprachverarbeitung schwierig?

Die maschinelle Sprachverarbeitung hat mittlerweile eine lange Tradition – lang für eine junge Wissenschaft wie die Informatik. Die ersten Arbeiten reichen in die 1960er Jahre zurück. Nach der anfänglichen Euphorie war der Fortschritt deutlich langsamer als erwartet. Woran lag das?

Bei der Verarbeitung von Sprache stellte es sich nach und nach heraus, dass sich diese dem Zerlegen in einfach formulierbare Regeln und Operationen widersetzt. Das erscheint verblüffend, da jeder Mensch imstande ist, Sprache zu beherrschen und zu verstehen. Eine Analyse zeigt allerdings, dass dies zunächst einmal nur für die Muttersprache gilt. Am Vergleich mit einer Fremdsprache wird klar, dass es nicht selbstverständlich ist, eine beliebige Sprache zu erlernen und zu beherrschen.

Die Gründe, die es für Menschen schwierig machen, eine Sprache zu erlernen, gelten auch in ähnlicher Form für Computer: Bei dem Versuch, eine fremde Sprache zu erlernen, wird niemand zum kompetenten Sprecher und Hörer, indem er Vokabeln, Aussprache und grammatikalische Regeln kennt. Stattdessen ist es viel besser, praktische Erfahrungen beim Zuhören, Sprechen, Lesen und Übersetzen zu sammeln.

Dementsprechend geht auch der erfolgreiche Ansatz zur maschinellen Sprachverarbeitung davon aus, dass der Computer aus Beispiel-

daten lernt und mit Plausibilitätsbewertungen statt mit vorgegebenen kategorischen Regeln arbeitet. Die Grundlagen dafür bilden die statistische Entscheidungstheorie und das maschinelle Lernen, zu dem auch die neuronalen Netze zu rechnen sind.

Holistischer Ansatz und maschinelles Lernen

Der Lehrstuhl für Informatik 6 (Maschinelles Lernen und Sprachtechnologie) verfolgt einen Ansatz, den man als holistisch (= ganzheitlich) bezeichnen kann. Dabei werden nicht einzelne Laute oder Buchstaben isoliert verarbeitet, sondern ein Satz wird als eine Einheit betrachtet und in seiner Gesamtheit vom Computer erfasst. Hier sei zum besseren Verständnis wieder die Analogie zum menschlichen Spracherkennen herangezogen. Auch der menschliche Hörer gleicht Unzulänglichkeiten im Gehörten, die durch weggelassene Laute, dialektgefärbte Aussprache, Umgebungsgeräusche und so weiter bedingt sind, in der Regel unbewusst aus, indem er zusätzliches Wissen über die Grammatik und die Bedeutung des Gesagten verwendet. Angesichts der dargelegten Schwierigkeiten ist klar, dass das entscheidende Kriterium für den Entwurf eines derartigen Systems die Performanz oder Leistungsfähigkeit sein muss, wobei dieses Performanzkriterium mathematisch genau zu definieren ist. Damit liegt eine Aufgabenstellung vor, wie sie prototypisch in der statistischen Entscheidungstheorie untersucht wird. Die Problemstellung geht in ihrem Ansatz auf Arbeiten von Thomas Bayes im 18. Jahrhundert zurück, so dass die entsprechende Entscheidungsregel als „Bayessche Entscheidungsregel“ bezeichnet wird. Für die Sprachverarbeitung, konkret die Spracherkennung, geht der Einsatz der statistischen Entscheidungstheorie auf Arbeiten in den 1970er Jahren zurück. Um den Ansatz

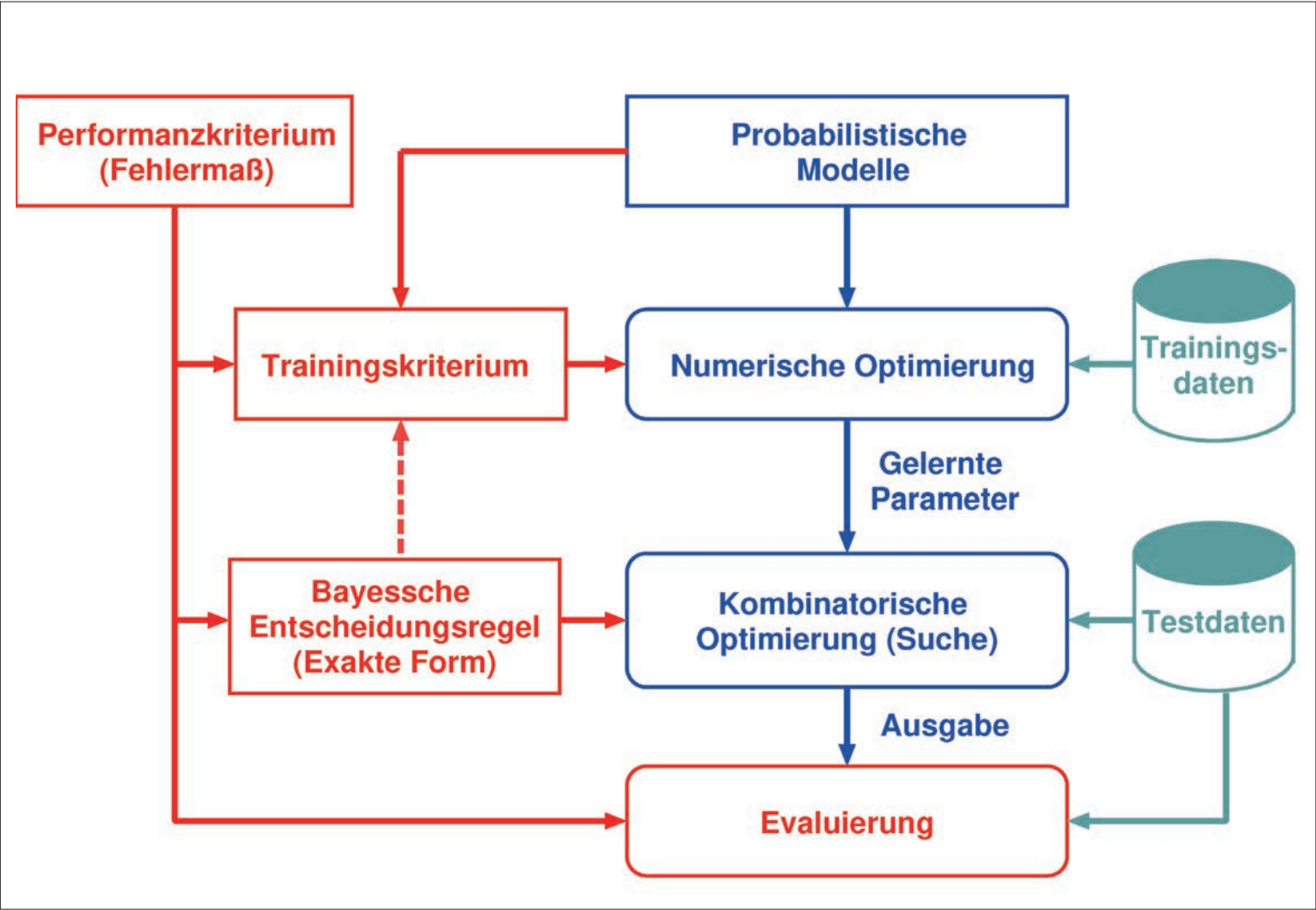


Bild 2: Holistischer Ansatz und maschinelles Lernen

mathematisch genauer zu beschreiben, wird auf die drei in Bild 1 dargestellten Aufgaben zurückgegriffen. Für jede der drei Aufgaben kann eine Eingangsfolge von Daten (für die Sprach- und Handschrifterkennung) oder von Symbolen (Buchstaben oder Wörter für die Übersetzung) und eine Ausgangsfolge von Wörtern definiert werden. Allgemein wird die Eingangsfolge als eine Folge von Elementen (Daten oder Symbolen) über der Positionsachse $t = 1, \dots, T$: beschrieben:

$$x_1^T := x_1, \dots, x_t, \dots, x_T,$$

Bei der Ausgangsfolge handelt es sich für jede der drei Aufgaben um eine Wortfolge unbekannter Länge N mit den Wortpositionen $n = 1, \dots, N$:

$$c_1^N := c_1, \dots, c_n, \dots, c_N.$$

Die Bayessche Entscheidungsregel geht davon aus, dass der Zusammenhang zwischen der beobachteten Eingangsfolge x_1^T und

der zu generierenden Ausgangsfolge c_1^N durch eine Posterior-Verbundverteilung $p(N, c_1^N | x_1^T)$ erfasst wird, die prinzipiell in verschiedenen Formen realisiert werden kann: als eine große überdimensionale Tabelle oder in Form eines Algorithmus, der diese Wahrscheinlichkeiten bei Bedarf produzieren kann. Unter vereinfachenden Annahmen lautet die Bayessche Entscheidungsregel, mit der die Folge der unbekannten Ausgangswörter generiert wird:

$$x_1^T \rightarrow \hat{c}_1^N(x_1^T) := \arg \max_{(N, c_1^N)} p(N, c_1^N | x_1^T)$$

Das bisher Beschriebene ist das Konzept des holistischen Ansatzes in einer einfachen Form. Um ein reales Sprachverarbeitungssystem mit diesem Ansatz aufzubauen, sind viele Einzelheiten auszuarbeiten und weitere Fragestellungen zu untersuchen. An dieser Forschungsthematik wird am Lehrstuhl für Informatik 6 (Maschinelles Lernen und Sprachtechnologie)

seit mehr als 25 Jahren geforscht. Diese Arbeiten wurden und werden im Rahmen von Verbundprojekten auf europäischer (TC-Star, EU-Bridge, ERC) und internationaler Ebene (US-ARPA-Projekte GALE, BOLT, BABEL und Google) durchgeführt.

Bild 2 gibt einen Überblick über die Architektur des holistischen Ansatzes. Dargestellt sind die Komponenten eines Sprachverarbeitungssystems, wobei auch versucht wird, die wechselseitigen Abhängigkeiten (zumindest näherungsweise) zu erfassen. Zu den Komponenten im Einzelnen:

Probabilistische Modelle

Die Modellierung der Posterior-Verbundverteilung $p(N, c_1^N | x_1^T)$ ist die wichtigste Komponente eines datengetriebenen Ansatzes, da diese Verteilung in direkter Wechselwirkung mit allen anderen Komponenten des Systems steht. Die wissenschaftliche Heraus-

forderung besteht darin, die Abhängigkeiten sowohl innerhalb der Eingangsfolge x_1^T und der Ausgangsfolge c_1^N als auch zwischen diesen beiden Folgen in geeigneter Weise mit mathematischen Strukturen zu erfassen. Die konkrete Modellierung und Strukturierung hängt von der Aufgabenstellung (Spracherkennung, Handschrifterkennung, Übersetzung) ab und kann immer nur eine Approximation an die wahren Abhängigkeiten liefern. Als Beispiele für mathematische Strukturen seien genannt: hidden Markov models, recurrent neural networks, connectionist temporal classification, attention models.

Performanzkriterium

Das Performanzkriterium soll quantitativ erfassen, wie gut das System als Ganzes die Sprachverarbeitungsaufgabe erfüllt. Wenn man die einzelnen Komponenten des Gesamtsystems entwirft, ist deswegen streng genommen immer die Frage zu stellen: Wie ist der Zusammenhang mit der Performanz des Gesamtsystems? Ein Beispiel eines Performanzkriteriums ist die Zählung von Wortfehlern, wobei allerdings die Wortstellung, insbesondere in der Übersetzung, in geeigneter Weise zu berücksichtigen ist. Als Verfeinerung kann man Wortfehler gewichten, um die spezielle Relevanz einzelner Wörter für die konkrete Anwendung zu berücksichtigen.

Bayessche Entscheidungsregel

Die exakte Form der Bayesschen Entscheidungsregel hängt vom Performanzkriterium ab und sollte – strenggenommen – eine Auswirkung auf das Trainingskriterium und die Generierung des Ausgangssatzes haben. Eine einfache Form der Bayesschen Entscheidungsregel wurde oben genannt. Diese Form der Entscheidungsregel wird in vielen Systemen als Näherung verwendet.

Trainingskriterium

Die Verbundverteilung (oder Modell) $p(N, c_1^N | x_1^T)$ hat heutzutage hunderte Millionen von freien Parametern, beispielsweise die Gewichte eines neuronalen Netzes. Im Idealfall sorgt das Trainingskriterium dafür, dass diese freien Parameter so trainiert werden, dass das resultierende Modell möglichst gut im Sinne des Performanzkriteriums ist. Mit anderen Worten stellt sich die Frage nach dem mathematischen Zusammenhang zwischen Performanzkriterium und Trainingskriterium.

Numerische Optimierung

Das Trainingskriterium definiert ein mathematisches Optimierungsproblem, für das in der Regel keine geschlossene Lösung mehr möglich ist. Deswegen braucht man Verfahren der numerischen Optimierung, um die optimalen Parameter zu finden. Die Herausforderung liegt darin, dass die heutigen Modelle eine riesige Zahl von freien Parametern haben. Als Beispiel seien hier neuronale Netze genannt, bei denen in der Regel cross-entropy als Trainingskriterium und verschiedene Varianten von Gradientenverfahren (backpropagation) verwendet werden.

Trainingsdaten

Das Trainingskriterium wird anhand von annotierten Daten formuliert. Im Beispiel der in Bild 1 dargestellten Aufgabenstellungen sind dies Paare von Eingangs- und Ausgangsfolgen. Der Umfang dieser Paare kann sehr groß sein: 1000 Stunden annotierte (oder transkribierte) Audio-Daten für die Spracherkennung und 100 Millionen Satzpaare von Quell- und Zielsprache für die Sprachübersetzung. Im Idealfall sind diese Trainingsdaten repräsentativ für die tatsächliche Anwendung des Sprachverarbeitungssystems. Einen Spezialfall von Trainingsdaten stellen reine Textdaten in natürlicher Sprache dar. Derartige Daten liefern Beispiele für die Ausgangsfolgen des Sprachverarbeitungssystems (ohne dass eine Eingangsfolge vorliegt). Der Vorteil dieser reinen Textdaten ist, dass man typischerweise hunderte von Millionen von Beispielsätzen finden kann (praktisch ohne Kosten) und man daraus ein linguistisches probabilistisches Sprachmodell trainieren kann, das Syntax, Semantik und Pragmatik der Sprachverarbeitungsaufgabe approximativ modelliert.

**Kombinatorische Optimierung
(Suche, Generierung)**

Die oben gezeigte einfache Form der Bayesschen Entscheidungsregel erfordert eine Bestimmung der Wortfolge mit der größten Wahrscheinlichkeit. Dies ist ein kombinatorisches Optimierungsproblem, bei dem im Prinzip alle möglichen Wortfolgen betrachtet werden müssen. Da die Zahl der möglichen Wortfolgen astronomisch groß ist (Beispiel: ca. 10^{100} bei einem Vokabular von 100.000 Wörtern und einer Satzlänge von $N = 20$ Wörtern), braucht man effiziente und leistungsfähige Algorithmen, um den Ausgabesatz zu generieren. In Abhängigkeit vom konkreten Performanzkriterium wird die exakte Form der Bayesschen Entscheidungsregel nochmals

komplizierter sein, so dass geeignete Approximationen benötigt werden.

Testdaten und Evaluierung

Um die Leistungsfähigkeit eines Sprachverarbeitungssystems möglichst realistisch zu messen, führt man eine Evaluierung an Testdaten durch. Die Anforderungen an die Testdaten sind: Sie sollen repräsentativ für die praktische Anwendung und vollkommen unabhängig von den Trainingsdaten sein. Anhand dieser Testdaten kann man die Leistungsfähigkeit des Systems messen und die Fehler und Schwachstellen analysieren. Aus dieser Analyse können sich Rückschlüsse auf die Schwachstellen des Systems ergeben, die dann entsprechend zu verbessern sind. Insgesamt gesehen läuft der holistische Ansatz auf ein extrem komplexes mathematisches Optimierungsproblem hinaus. Wegen dieser Komplexität muss man, wie oben dargestellt, das Problem in einzelne Komponenten zerlegen und geeignete Approximationen einführen, wobei immer die Auswirkung auf die Performanz des Gesamtsystems im Vordergrund stehen sollte.

Autor
Univ.-Prof. Dr.-Ing. Hermann Ney war bis 2020 Inhaber des Lehrstuhls für Informatik 6 (Maschinelles Lernen und Sprachtechnologie).

Fehler in der Steuerungssoftware – eine wissenschaftliche Herausforderung

Technische Systeme werden sicherer mit Informatik-Methoden

Control software in technical applications, called “embedded software”, is ubiquitous. It determines the functionality of virtually any system we get in touch with in our daily life, ranging for example from household appliances and cars to medical equipment and industrial automation. Obviously, errors in embedded software can lead to severe damage and injuries or even deaths of people. The article presents recent methods from computer science for detecting or avoiding such “bugs”. The main approach is based on so-called “formal methods”. This means that we use mathematics to describe the intended behavior of an embedded program as well as its possible behavior, and we check that the latter conforms to the former with the help of computer algorithms. Our particular interest lies in making such methods applicable to real-world problems by simplifications or over-estimations. We report on successful examples in industrial automation, partly from the Excellence Cluster “Internet of Production”. Further success stories refer to spin-off companies which provide technologies for improving industrial testing and for enhancing standard mobile operating systems with industrial robustness and real-time capabilities.



Noch vor wenigen Jahrzehnten kam man im Alltag nur dann mit Computerprogrammen in Berührung, wenn man sich vor einem Bildschirm befand und versuchte, über eine Tastatur den Rechner dazu zu bewegen, das zu tun, was man von ihm wollte. Diese Situation hat sich grundlegend geändert. Heute vergeht kaum eine Minute, in der kein software-gesteuertes System unser Leben beeinflusst.



Bild 1: Bei Medizingeräten ist es besonders offensichtlich, dass Fehler in der Steuerungssoftware unbedingt vermieden werden müssen. Daher kommt ein Teil der in diesem Artikel beschriebenen Methoden auch bei der Entwicklung des Notfall-Beatmungsgeräts PV1000 für Krisensituationen (<http://www.pv1000.de>) zum Einsatz. Das Bild zeigt das Gerät im simulierten Einsatz und die Projektinitiatoren Professor Lutz Eckstein, Professor Steffen Leonhardt, Professor Stefan Kowalewski (stehend von links) sowie Dr. André Stollenwerk und Dr. Marian Walter (sitzend von links).
Foto: Peter Winandy

Software bremst und beschleunigt unsere Autos, sie öffnet die Türen der Aufzüge, sie beatmet kranke Patienten und sie steuert Produktionsanlagen. Mittlerweile vernetzt sie auch die meisten dieser Systeme und sorgt zum Beispiel dafür, dass in einer Fertigung automatisch ein anderer Auftrag vorgezogen wird, wenn der Lastwagen mit der eigentlich benötigten Lieferung im Stau steht.

Fehler in der Steuerungssoftware für technische Systeme – man spricht hier von „eingebetteter (englisch embedded) Software“ – können besonders schwerwiegende Folgen haben. Nicht selten führen sie zu hohen wirtschaftlichen Schäden, zum Beispiel, wenn durch Softwareprobleme Fabriken für Stunden stillgelegt werden, neue PKW-Modellreihen monatelang nicht auslieferungsfähig

sind oder ein Testflug scheitert. Noch schlimmer ist es natürlich, wenn Menschen ums Leben kommen, weil Software versagt. Das Finden, Beseitigen oder auch Vermeiden von Fehlern in Software ist seit jeher ein Kernthema in der Informatik. Der Lehrstuhl für Informatik 11 (Embedded Software) hat sich darauf spezialisiert, dies für technische Steuerungssoftware zu erforschen. Die

wissenschaftlich weitreichendsten Ansätze dazu verwenden „formale“, das heißt mathematisch-logische Methoden. Sie stammen größtenteils aus der theoretischen Informatik, die sich damit beschäftigt, grundlegende Eigenschaften und Grenzen dieser Ansätze zu erkennen und zu beweisen. Die Forschung am Lehrstuhl für Informatik 11, der zur praktischen Informatik gehört, hat dagegen das Ziel, formale Methoden für die Industrie anwendbar zu machen.

Grundidee der formalen Verifikation von Steuerungssoftware ist die folgende: Wenn man eine Beschreibung sowohl des möglichen als auch des gewünschten Verhaltens eines technischen Systems hat, kann man – auch automatisch mit dem Computer – analysieren, ob nicht gewünschtes Verhalten überhaupt möglich ist. Trifft dies zu, liegt ein Fehler in der Steuerungssoftware vor. Die Beschreibung des möglichen Verhaltens, das so genannte Modell, kann zum Beispiel für einen Roboter enthalten, welche Positionen er überhaupt anfahren kann. Das Modell des gewünschten Verhaltens wird bestimmte Positionen ausschließen, weil sich dort möglicherweise ein Mensch aufhält. Die Verifikation muss dann zeigen, dass nur erlaubte Positionen möglich sind.

Hoher Rechenaufwand und Speicherbedarf

Das praktische Problem der formalen Verifikation ist der Aufwand für die Erstellung der Verhaltensmodelle und für ihre Analyse. Für Systeme in realistischer Größenordnung benötigen die grundlegenden Algorithmen zu viel Zeit, um zu einem Ergebnis zu kommen,

oder mehr Speicherplatz, als der größte Rechner der Welt zur Verfügung hätte. Die Forschung in der praktischen Informatik befasst sich daher vor allem damit, mit welchen Vereinfachungen oder Näherungslösungen man Lösungsmethoden gestalten kann, die für die industrielle Anwendung praktikabel sind.

Ein Beispiel dafür ist die Wertemengenanalyse für Programme von Speicherprogrammierbaren Steuerungen, kurz SPS, die in der Industrie am häufigsten anzutreffenden Automatisierungsgeräte. Hierbei wird eine Technik eingesetzt, die als „statische Analyse“ bezeichnet wird. „Statisch“ bedeutet, dass das Programm weder ausgeführt wird, noch alle möglichen Abläufe exakt bestimmt werden. Vielmehr wird durch Abschätzungen auf der Basis des Programmtextes versucht, wichtige Eigenschaften des Softwareverhaltens zu ermitteln. Dabei ist es allerdings möglich, dass Hinweise auf ein mögliches Fehlverhalten nur dadurch entstehen, dass die Abschätzung zu grob ist. In Wirklichkeit liegt dann also kein Fehler vor. Ergibt die Analyse allerdings, dass keine ungewünschten Eigenschaften vorhanden sind, kann man sich auf das Ergebnis verlassen.

Am Lehrstuhl für Informatik 11 wurde mit dieser Technik ein Werkzeug entwickelt, das mit geringem Aufwand alle Werte bestimmt, die eine Variable in einem SPS-Programm in jedem überhaupt möglichen Programmablauf annehmen kann. Die Anwendung auf industrielle Beispiele zeigt, dass diese Art der Analyse hilfreiche Hinweise auf mögliche Software-Bugs gibt. So wurde zum Beispiel ein Fehler in einer international veröffentlich-

ten, standardisierten Sicherheitssteuerung gefunden, weil eine Variable zur Beschreibung unterschiedlicher Programmmustände einen nicht vorgesehenen Wert annehmen konnte.

Änderung der Steuerungssoftware im Betrieb

Eine noch größere Bedeutung gewinnen Verfahren zum schnellen Auffinden von Fehlern, wenn Maschinensteuerungen im Betrieb verändert werden. Ein solches Szenario wird im Exzellenzcluster „Internet der Produktion“ untersucht, in dem zahlreiche Institute und Lehrstühle der RWTH gemeinsam an einer zukünftigen, datengesteuerten Produktionstechnik forschen. Das übergreifende Thema ist: Wie kann die Produktion durch Vernetzung und Auswertung der Riesenmengen an anfallenden Daten effizienter, zuverlässiger und sicherer gemacht werden? Insbesondere wird der Einsatz von künstlicher Intelligenz und maschinellen Lernverfahren untersucht, um aus erfassten Daten Möglichkeiten zur Verbesserung von Produktionsprozessen abzuleiten. Wenn man die entsprechenden Abläufe auf der Basis dieser Erkenntnisse anpasst, muss man auch die Steuerungssoftware ändern. Daraus ergibt sich sofort die Notwendigkeit, schnell und ohne langen Stillstand der Anlage sicherzustellen, dass durch diese Änderungen nicht ungewollt neue Fehler in die Steuerungsprogramme gebracht werden.

In der Forschung wird untersucht, ob man Verfahren aus dem Engineering von neuen Anlagen auf dieses Szenario übertragen kann. Auch bei der Neuentwicklung von Pro-

duktionssystemen und ihrer Steuerungsprogramme gibt es nämlich Änderungen in Form von Fortschritten bei der Entwicklung. In jedem Schritt muss geprüft werden, ob neu hinzugekommene Funktionen nicht zu neuen Fehlern in der schon bestehenden Software geführt haben. Der Gedanke liegt nahe, diese so genannte „Regressionsverifikation“ auch auf die Situation anzuwenden, in der bereits laufende Steuerungssoftware zur Verbesserung der Produktion geändert wird. Die Herausforderung dabei ist, mit der Prüfung nicht von vorne anfangen zu müssen, sondern möglichst viele Prüfergebnisse über den nicht geänderten Teil der Steuerungssoftware wiederzuverwenden. In dem Exzellenzcluster wird dieser Ansatz an einer flexiblen Produktionsanlage im Werkzeugmaschinenlabor weiterentwickelt.

Formale Methoden unterstützen Testprozesse

Neben den beschriebenen formalen Methoden hat aber auch das klassische Testen seine Bedeutung für das Erkennen und Beseitigen von Softwarefehlern keineswegs verloren. Im Gegenteil: Obwohl Testen den Nachteil hat, dass damit nur stichprobenartig nach Fehlern gesucht werden kann, ist es weiterhin die in der Praxis mit Abstand am häufigsten verwendete Methode. Wissenschaftlich interessant ist daher die Frage, wie man den Testprozess noch besser unterstützen kann. Ein Ansatzpunkt sind geeignete Sprachen, um die auszuführenden Testfälle besonders verständlich beschreiben und auch automatisch ablaufen lassen zu können. Sie helfen am meisten, wenn die Fehlersu-

che damit schon in sehr frühen Entwicklungsphasen durchgeführt wird, in denen die Steuerungsprogramme noch nicht als Programmcode sondern auch hier als „Modelle“ bezeichnete Beschreibungen vorliegen. Dieser Ansatz wird mittlerweile von einer Ausgründung aus dem Lehrstuhl für Informatik 11, der Mindmotiv GmbH, unter anderem in Zusammenarbeit mit einem Automobilhersteller eingesetzt und weiterentwickelt.

Vermeidung von Fehlern im zeitlichen Verhalten von Software

Eine andere Möglichkeit, die Qualität von Steuerungssoftware zu erhöhen, ist es, Ausführungsplattformen bereitzustellen, die die Vermeidung von Fehlern unterstützen. Eine spezielle Fehlerquelle sind Verletzungen von so genannten Echtzeitanforderungen. Damit meint man, dass Programme nicht rechtzeitig ihr Ergebnis berechnen. Ein typisches Beispiel ist die Steuerung eines Airbags: Wenn der entsprechende Algorithmus zu lange braucht, um zu erkennen, dass ein Aufprall passiert, und den Airbag zu spät auslöst, sind schlimme Verletzungen die Folge. Echtzeitprobleme treten meistens auf, wenn mehrere Teilprogramme auf einem Computer laufen und ein Programm zu lange auf das andere warten muss. Um dies zu vermeiden, benötigt man ein Betriebssystem, das Hilfsmittel bereitstellt, mit denen das Zeitverhalten der Programme übersichtlich und nachvollziehbar spezifiziert werden kann. Das von Dr. Igor Kalkov, einem ehemaligen Doktoranden des Lehrstuhls für Informatik 11, gegründete Unternehmen Emteria GmbH hat das im mobilen Bereich beliebte Betriebssystem

Android um solche und andere Hilfsmittel erweitert. So können Steuerungsprogramme jetzt als Apps auf entsprechende Geräte geladen werden und industrielle Steuerungsaufgaben übernehmen.

Autor

Univ.-Prof. Dr.-Ing. Stefan Kowalewski ist Inhaber des Lehrstuhls für Informatik 11 (Embedded Software).

Alle Wege durch das Labyrinth

Die automatisierte Suche nach Softwarefehlern

Nuclear power plants, smartphones, and automatic doors have something in common: They do not work as intended if the computer software that controls them misbehaves. The current state of the art to ensure that software behaves correctly is to invest effort – and money – into creating test cases for as many situations as the developers can think of. However, not only will the real-world always find a spanner to throw in the works, but attackers will also attempt to cause situations that the original programmers did not consider.

One possible solution for this problem is automated testing. At first glance, this sounds like the perfect solution, as computers do not tire and will faithfully test millions of configurations. But computers are also uncreative and will faithfully test millions of equivalent configurations instead of recognizing and exploiting commonalities. The symbolic execution technique recognizes both the strengths and weaknesses of automated testing by teaching computers advanced reasoning, so they can be smart about exhaustive testing.

At the Chair of Communication and Distributed Systems, we not only use symbolic execution to analyze distributed applications and protocols, but also research different ways to improve the reasoning done by symbolic execution.

Software ist allgegenwärtig. Egal, ob es sich um einen Kernreaktor, ein Smartphone oder den Taster zum Öffnen der Bustüren handelt, wir alle müssen uns ständig darauf verlassen, dass Software sich korrekt verhält. Gleichzeitig sind Softwarefehler mit mehr oder minder tragischen Konsequenzen regelmäßig in den Nachrichten. Von digitalen Assistenten, die ihre Besitzer nicht verstehen über das neueste Daten-Leak bis hin zu Flugzeugabstürzen ist ein breites Spektrum von Pannen bis Katastrophen abgedeckt.

Bisher gibt es kein Allheilmittel für Softwarefehler, sodass sich nach wie vor Softwareentwickler selbst auf die Suche nach Fehlern machen müssen. Dies macht Softwarequalität jedoch zu einer unternehmerischen Entscheidung: Je mehr Aufwand – und damit Geld – in die Suche nach Fehlern investiert wird, desto weniger Fehler treten wahrscheinlich später auf. Für Unternehmen, die sicherheitskritische Anwendungen wie Flugzeugcomputer herstellen, kann die Fehlersuche so deutlich teurer werden als die ursprüngliche Entwicklung.

Neben den Kosten führt langwieriges manuelles Testen von Software auch zu einem Ermüdungseffekt bei den damit betrauten Entwicklern. Selbst für einfachste Programme müssen die Entwickler den Überblick über Unmengen von möglichen Situationen behalten. Zum Beispiel kann eine Minute

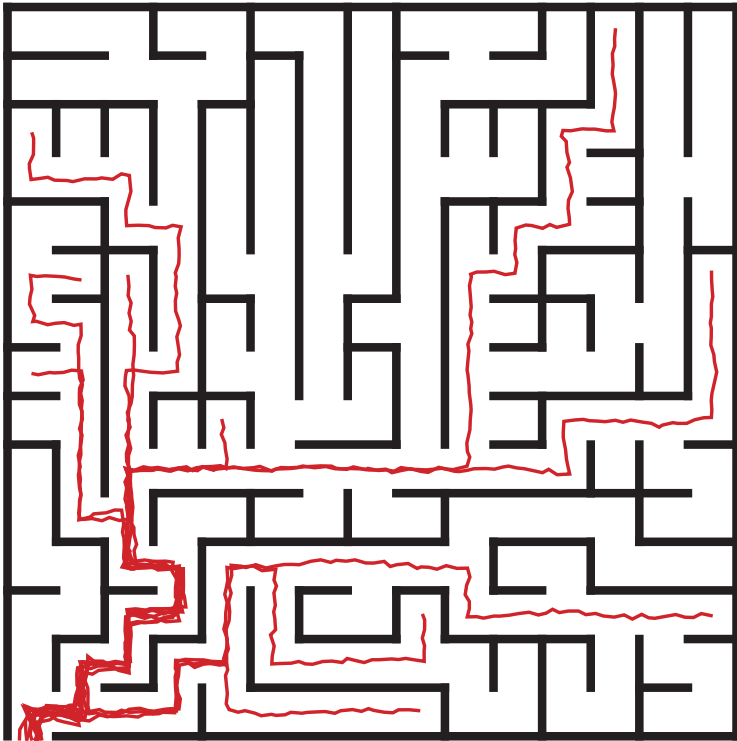


Bild 1: Software als ein Labyrinth von Entscheidungen. Menschen testen zwar effizient, aber nicht viel.

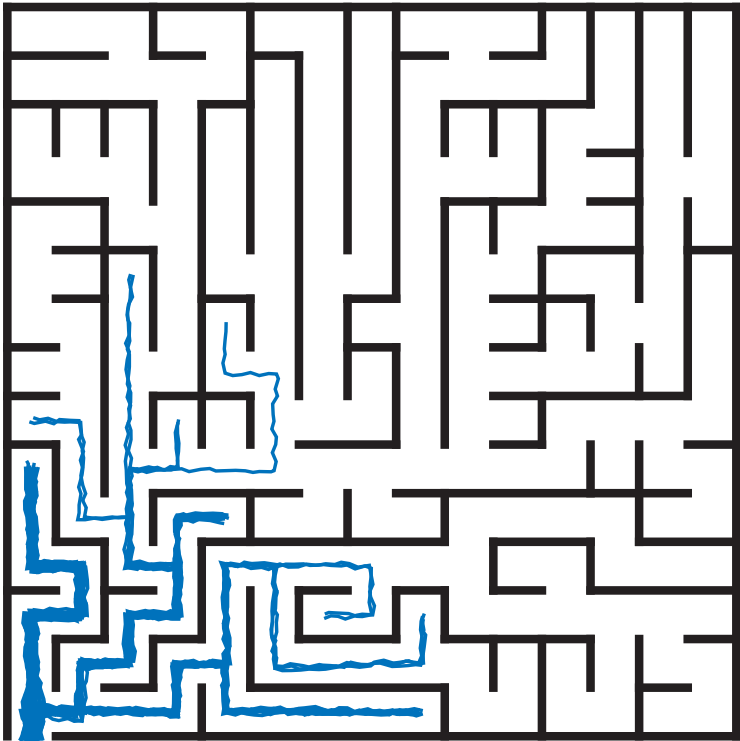


Bild 2: Einfache automatisierte Testmethoden können viele Pfade durchtesten, arbeiten aber meist zufällig und damit nicht sehr effizient.

aus 60 oder 61 Sekunden bestehen (wenn sie eine Schaltsekunde beinhaltet) und der Buchstabe „ä“ ungleich dem Buchstaben „ä“ sein – beispielsweise in der einen Form direkt als „ä“ gespeichert und in der anderen als ein „a“ und zusätzlich die Umlautpunkte „““. Jede noch so unwahrscheinliche Fehlersituation kann fatal sein – sei es einfach durch die millionenfache Verbreitung mancher Software oder durch aktive Angreifer, die eine solche Situation bewusst herbeiführen.

In den letzten Jahren feierte mit „Fuzzing“ ein fundamental anderer Ansatz sein Debut. Hierbei werden die von den Entwicklern vorgegebenen Testfälle zusätzlich von dem Computer verändert. Auf diese Weise kann aus einem einzelnen Testfall eine ganze Serie werden. Der Computer übernimmt den Teil der Arbeit, die für Menschen besonders schwierig ist, indem er stupide Millionen von Testfällen generiert. Da aber keine wirkliche Intelligenz im Spiel ist, verfährt Fuzzing ziellos und es bleibt unklar, wie gut das Programm wirklich getestet ist.

Symbolische Ausführung

Am Lehrstuhl für Informatik 4 (Kommunikation und verteilte Systeme) wird an der Technik „Symbolische Ausführung“ geforscht, die Computern beibringt, Testfälle wie ein Mensch zu erzeugen. Hierzu ist es hilfreich, sich ein Computerprogramm als eine lange Reihe von

Wenn-Dann-Abfragen vorzustellen: Wenn die Datei existiert, dann öffne sie. Wenn die geöffnete Datei das richtige Format hat, dann lade den Text. Wenn eine Seite voll ist, dann erhöhe die Seitenanzahl.

Um ein Programm nun vollständig zu testen, muss jede mögliche Kombination der Wenn-Dann-Abfragen getestet werden. Eine nützliche Analogie ist hier die Erforschung eines Labyrinths. Die Abzweigungen stehen hier für die Wenn-Dann-Abfragen, und am Ende sollen alle Schätze, die irgendwo in dem Labyrinth versteckt sind, geborgen sein.

Menschliche Tester sind vergleichsweise teuer und ermüden schnell, sodass wie in Bild 1 nur wenige Pfade durch das Labyrinth erforscht werden. Einfache computergestützte Methoden probieren zwar ermüdungsfrei und relativ kostengünstig sehr viele verschiedene Pfade, tendieren aber dazu, immer wieder ähnliche Fälle zu erwischen. Wie in Bild 2 ersichtlich, führt dies zu einer Vervielfachung des Aufwandes. Wenn man das Labyrinth jedoch mit symbolischer Ausführung erforscht, wird, wie in Bild 3, jeder Pfad nur genau einmal getestet, ohne dass ein Mensch die Fälle vorgibt.

Symbolische Ausführung beruht im Kern auf dem Erfüllbarkeitsproblem. Dieses fragt, ob es eine Menge von Zahlen gibt, sodass eine Gruppe von Formeln aufgeht. Für die Formel „x ist größer als fünf“ gibt es zum Beispiel

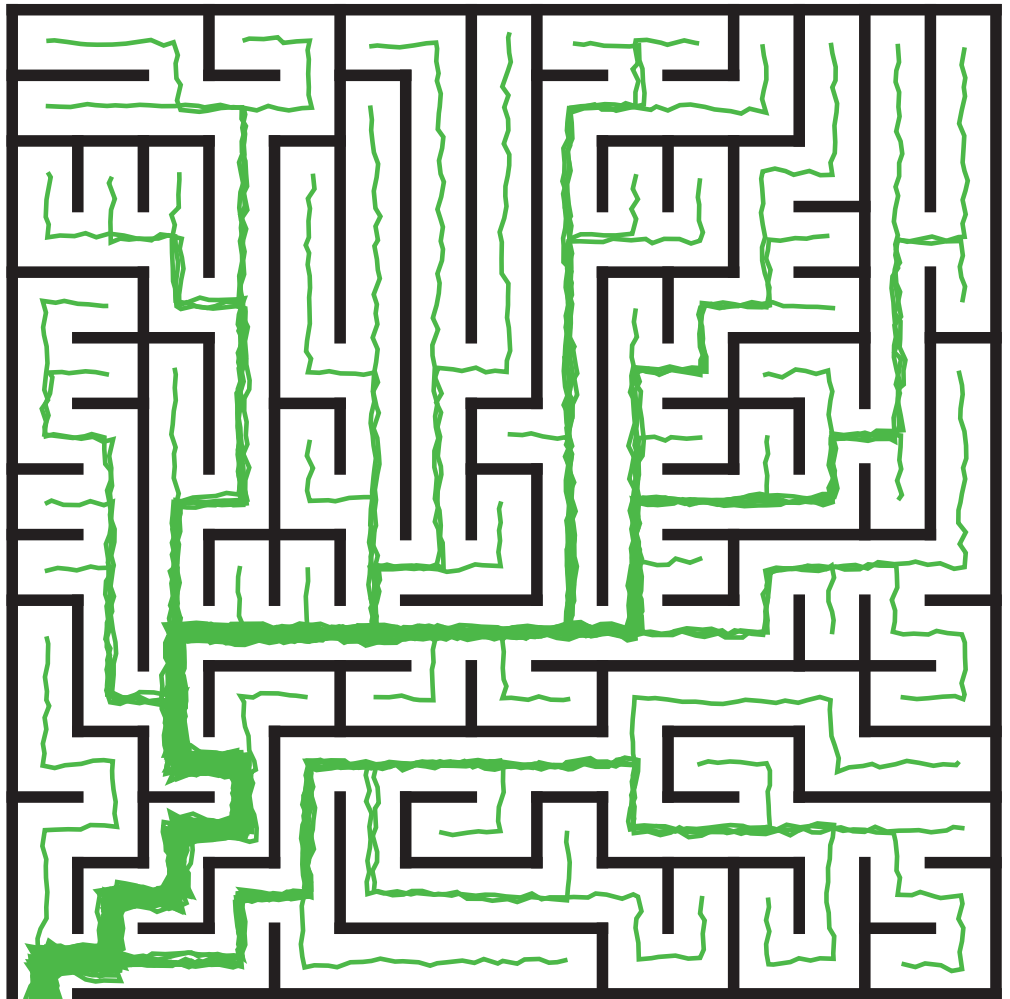


Bild 3: Symbolische Ausführung kann ein Programm mit der Geschwindigkeit und Beharrlichkeit eines Computers und trotzdem so effizient wie ein Mensch testen.

viele Zahlen, die man für x einsetzen kann, sodass die Formel aufgeht, nämlich die Zahlen sechs, sieben, acht und so weiter. Eine Wenn-Dann-Abfrage, in der genau diese Frage gestellt wird, kann also von dem Programm erfüllt werden. Angenommen das Programm trifft jedoch später auf die Wenn-Dann-Abfrage „ x ist gleich null“, ergibt sich ein anderes Bild, da es keine Zahl gibt, die gleichzeitig größer als fünf und gleich null ist. Die Wenn-Dann-Abfragen sind also in dieser Kombination nicht erfüllbar.

Um das Erfüllbarkeitsproblem zu lösen wird jedoch sehr viel Rechenleistung benötigt. Erst in den letzten 20 Jahren ist die Forschung in diesem Feld weit genug vorangekommen, dass es in vielen Fällen praktisch lösbar geworden ist. Trotzdem ist symbolische Ausführung weiterhin häufig von der Erfüllbarkeitsüberprüfung limitiert. In der Praxis bedeutet dies, dass einfachere Techniken wie Fuzzing mehr Pfade ausprobieren können. Auch wenn dies die Abdeckung erhöht, sind ein-

fachere Techniken nur bedingt geeignet, ein Programm komplett abzudecken. Bild 4 zeigt zehnmal mehr zufällige Pfade als die Bilder 2 und 3, ohne jedoch das Labyrinth vollständig zu erforschen.

Am Lehrstuhl für Informatik 4 (Kommunikation und verteilte Systeme) konnte gezeigt werden, dass die Performance von symbolischer Ausführung nicht nur von den besonders schwierigen Anfragen begrenzt wird, sondern auch von der großen Menge an relativ einfach lösbaren. Eine besonders häufig auftretende Gruppe von Anfragen führte zu dieser Schlussfolgerung. Anfragen wie „ x ist größer als fünf und x ist kleiner als sieben“ ergeben sich an vielen Stellen eines Programms (zum Beispiel um zu testen ob ein Array-Zugriff valide ist) und sind einfach zu lösen. Daher wurde eine Komponente vor die eigentliche Erfüllbarkeitsüberprüfung geschaltet, die derartige Anfragen sehr schnell erkennt und beantwortet. Wenn die Anfrage nicht dem Muster entspricht, wird sie ohne Zeitverzug

an die Erfüllbarkeitsüberprüfung weitergeleitet. Damit konnte gezeigt werden, dass solche einfachen Erfüllbarkeitsabfragen einen großen Teil der Anfragen ausmachen – und dass sich hier auch kleine Verbesserungen in der Summe deutlich bemerkbar machen.

Symbolische Ausführung von verteilten Systemen

Interessant ist die Anwendbarkeit von symbolischer Ausführung auf verteilte Anwendungen und Protokolle. Die Nebenläufigkeit solcher Anwendungen erschwert allerdings die Analyse erheblich, da nicht mehr nur die Eingaben das Verhalten des Programms entscheiden, sondern auch, wann genau die einzelnen Anweisungen ausgeführt werden. Ein klassischer Fehler in solchen nebenläufigen Programmen ist, dass ein Teil des Programms bereits versucht, ein Ergebnis zu lesen, ohne sicherzustellen, dass es auch schon berechnet wurde. Wenn die beiden Programmteile in der „richtigen“ Reihenfolge ausgeführt

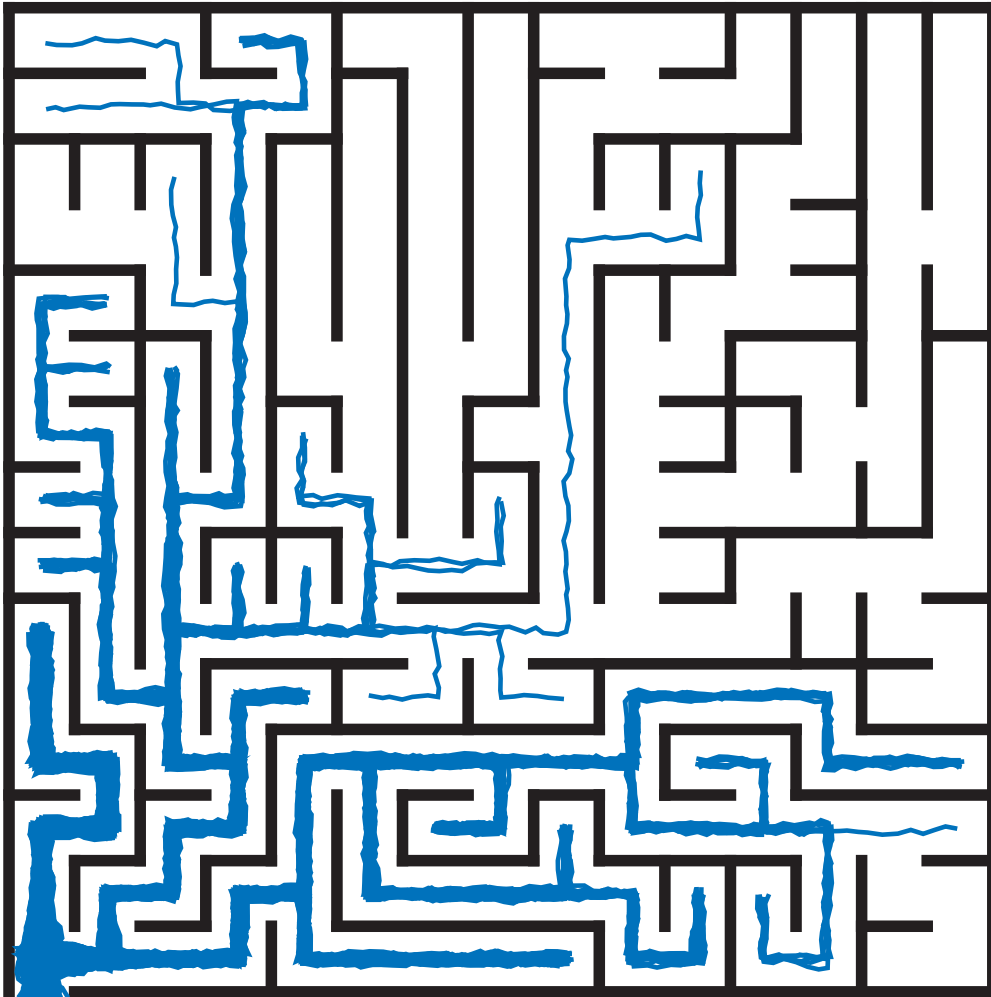


Bild 4: Auch zehnmal mehr Pfade führen bei stochastischen Methoden nicht unbedingt zu einer vollständigen Erforschung.

werden, läuft das Programm korrekt und alle Eingaben führen zu dem richtigen Ergebnis. Sollten sie aber in der „falschen“ Reihenfolge ausgeführt werden, entsteht ein Fehler. Auf den ersten Blick scheint die Lösung offensichtlich: Wenn die Reihenfolge wichtig für das Verhalten ist, müssen alle möglichen Reihenfolgen durchgetestet werden. Allerdings stellt sich leider heraus, dass es viel zu viele Möglichkeiten gibt, um diese Methode tatsächlich anzuwenden. Geforscht wird daher an mehreren Techniken, die alle das Problem auf eine grundsätzlich ähnliche Art und Weise angehen. Vorbild ist wieder der Mensch. Ein menschlicher Tester würde versuchen, die Stellen zu identifizieren, an denen eine Ressource von verschiedenen Programmteilen aus genutzt wird, denn nur dort kann die Reihenfolge einen Unterschied machen. Solange die nebenläufigen Programmteile nur unabhängig voneinander arbeiten, hat die Nebenläufigkeit an sich keinen Einfluss auf das Programmverhalten. Nun ergibt sich die

Frage, wie solche Stellen identifiziert werden können. Als einfache, aber effektive Lösung kann man sich auf Programme beschränken, in denen die Kommunikation zwischen den nebenläufigen Teilen explizit beschrieben ist. Die wichtigste Untergruppe bilden hier Programme, die für sich genommen nicht nebenläufig, aber in einem Netzwerk miteinander verbunden sind. Vernetzte Programme können einander nicht direkt beeinflussen, daher müssen also nur noch alle Reihenfolgen, in denen Netzwerkzugriffe passieren können, betrachtet werden. Darüber hinaus wird auch an einer Technik geforscht, die Programme, die mehrere CPU-Kerne auf einem Computer nutzen, analysieren kann, indem alle Speicherzugriffe auf parallele Zugriffe überprüft werden. Indem das Nebenläufigkeitsverhalten extrahiert wird, lässt sich die Menge der Varianten, die zu testen sind, minimieren.

Ein Blick in die Zukunft

Je mehr Software unsere Welt bestimmt, desto wichtiger wird es, dass Programme sich korrekt verhalten. Wir forschen an intelligenten Methoden, um die Software von morgen möglichst einfach und vollständig zu testen. Hierzu lassen wir uns davon inspirieren, wie menschliche Tester über Programme nachdenken und nutzen die Beharrlichkeit und Geschwindigkeit von Computern, um Software möglichst allumfassend zu testen.

Autoren

Daniel Schemmel, M. Sc., Julian Büning, M. Sc., und Felix Rath, M. Sc., sind wissenschaftliche Mitarbeiter am Lehrstuhl für Informatik 4 (Kommunikation und verteilte Systeme). Univ.-Prof. Dr.-Ing. Klaus Wehrle ist Inhaber des Lehrstuhls für Informatik 4 (Kommunikation und verteilte Systeme).



Bild 1: Impression aus dem Reinraum – Vereinzelung eines lithographisch strukturierten und gesägten Silizium-Wafers.

Foto: Peter Winandy

Energieeffiziente KI

Neuromorphe Hardware für künstliche intelligente Systeme

In the Information & Communication Technology (ICT) Profile Area, scientists from RWTH Aachen University, Forschungszentrum Jülich and the Johannes Rau Research Institute AMO work closely together on a revolution in computer architecture. Today's computers are lacking in cognitive capabilities that are comparable to those of the brain. In artificial intelligence (AI) applications such as pattern recognition tasks, they consume way too much energy. New "neuromorphic" computers, on the other hand, will combine processing principles of the brain and novel memristive switching devices to overcome the wall posed by the end of Moore's law. The novel hardware architectures and related algorithms provide the drive for autonomous and self-learning systems based on the operation and architecture of the highly efficient human brain.

Im Profilbereich „Information & Communication Technology“ (ICT) arbeiten Wissenschaftlerinnen und Wissenschaftler der RWTH Aachen, des Forschungszentrums Jülich und der AMO GmbH an einer Revolution in der Computerarchitektur. Für viele Anwendungen der Künstlichen Intelligenz (KI) sind die heutigen Computer nicht ausreichend leistungsfähig, sie verbrauchen bei Aufgaben der komplexen Mustererkennung zu viel Energie. Neuartige „neuromorphe“ Rechner versprechen eine deutlich verbesserte Energieeffizienz, weil sie sich an der Funktionsweise und der Architektur des hoch effizient arbeitenden menschlichen Gehirns orientieren. Der Energiehunger klassischer Rechner insbesondere für KI-Anwendungen liegt zu einem großen Teil in der Trennung von Prozessor- und Speichereinheit begründet. Eine solche klassische Rechnerarchitektur wird nach ihrem Erfinder „von Neumann“-Architek-

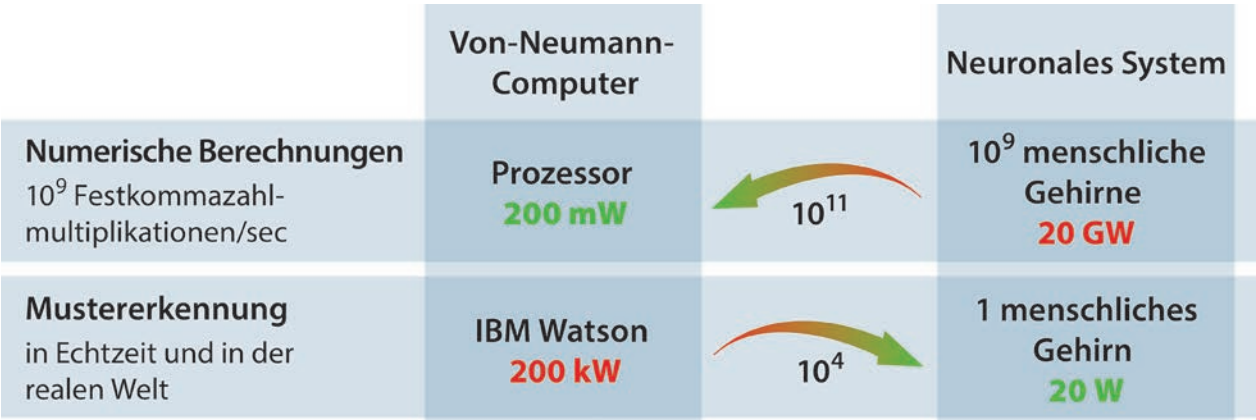


Bild 2: Energieverbrauch eines klassischen Rechners und des menschlichen Gehirns als biologisches neuronales Netz im Vergleich zwischen numerischen Berechnungen und kognitiven Aufgaben.



NUR Si-Oxidation
RWTH 2

NICHT FÜR



Bild 3: Vorbereitung zur Oxidation von Siliziumscheiben: Programmierung des Temperaturprofils und optische Inspektion
Foto: Peter Winandy

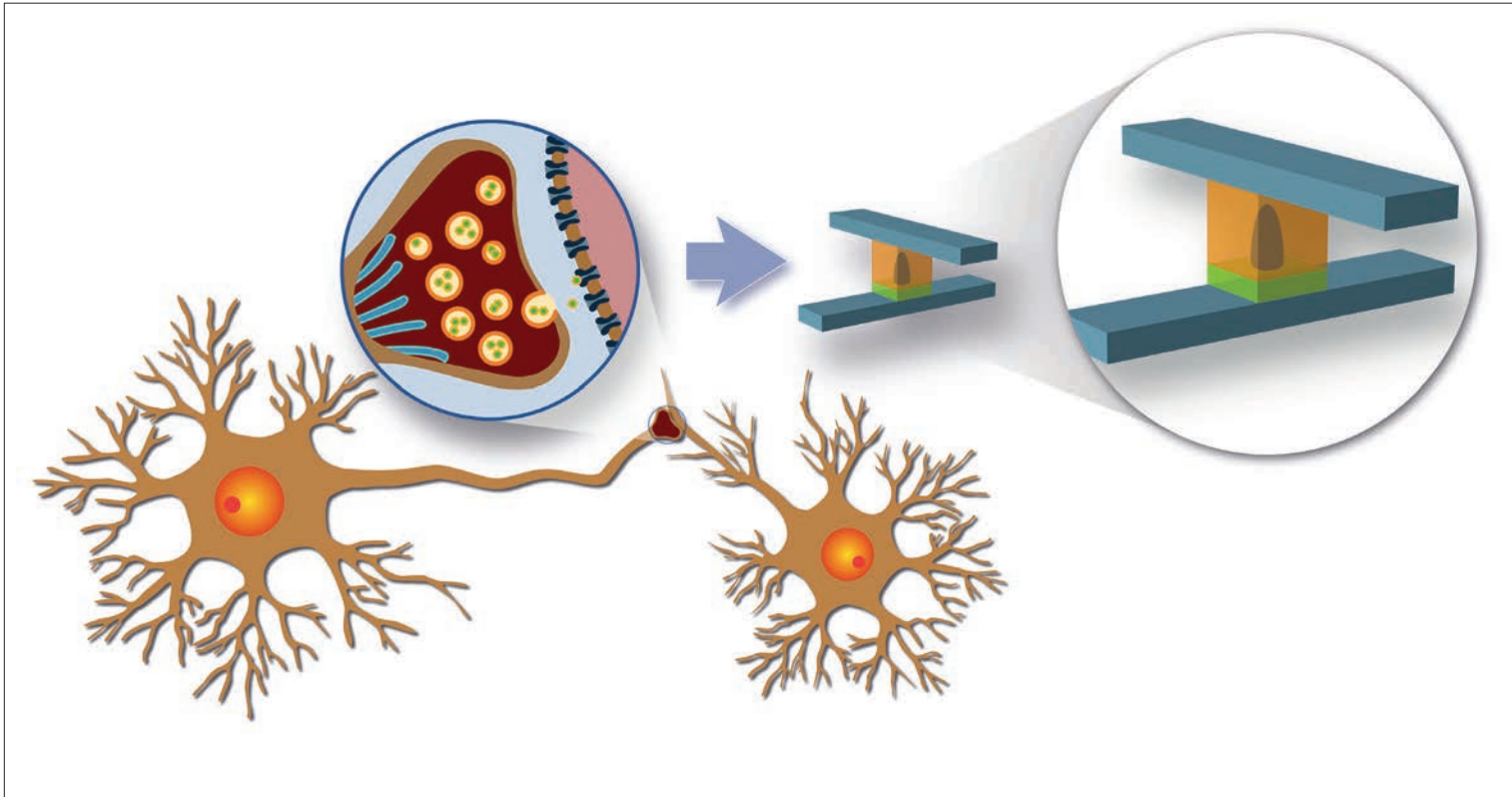


Bild 4: Schema des synaptischen Kontaktes zwischen zwei Nervenzellen (links). In der künstlichen Synapse (rechts) sitzen memristive Materialien zwischen zwei Leiterbahnen. Diese memristiven Materialien können ihren elektrischen Widerstand durch Spannungspulse aufgrund von atomaren Phasenwechsel- oder Valenzwechselprozessen ändern.

tur genannt. Für jede Rechenoperation werden Daten aus dem Speicher in den Prozessor geladen, dort werden dann die digitalen Rechenoperationen durchgeführt, und das Ergebnis wird wieder zurück in den Speicher geschrieben. Dieser ständige Verkehr zwischen Prozessor und Speicher führt zu einem Datenstau an diesem Flaschenhals, dem von Neumann-Bottleneck. KI-Leistungsfähigkeit und Energieeffizienz heutiger Systeme sind dadurch begrenzt.

In Bild 2 wird der Energieverbrauch eines klassischen Rechners mit dem des menschlichen Gehirns verglichen und tatsächlich kommt es dabei auf die Aufgabe an. Bei numerischen Berechnungen ist der von Neumann-Rechner dem Gehirn weit überlegen. Kognitive Anwendungen erledigt das Gehirn hingegen wesentlich energiesparender. Die Mustererkennung, die in der KI eine wesentlich größere Rolle einnehmen wird als numerische Berechnungen, erfordert den Umgang mit sehr großen Datenmengen, vielfach in Echtzeit oder schneller. Hierfür zeigt sich das menschliche Gehirn einem von Neumann-Rechner deutlich überlegen. Für Anwendungen in Bereichen der KI werden daher Rechner nach dem Vorbild des menschlichen Gehirns, sogenannte neuromorphe Rechnerarchitekturen, weltweit intensiv erforscht.

Angetrieben wird diese Forschung von der steigenden Nachfrage aus Wirtschaft und Gesellschaft. Die Analyse großer Datenmengen, Stichwort Big Data, ist ein zentraler Baustein kognitiver Funktionen, die alle KI-Anwendungen, wie zum Beispiel das Internet der Dinge, intelligente Städte, zukünftige Energienetze, autonomes Fahren, intelligente Produktionssysteme und personalisierte Medizin dominieren. Auch beim vernetzten Arbeiten ist es oftmals effizienter, intelligent vorverarbeitete Information auszutauschen statt Rohdaten.

Dieser gesellschaftliche Wunsch nach mehr intelligenter Rechenleistung bei gleichzeitigem Gebot eines nachhaltigen Umgangs mit Ressourcen wird von Seiten der weltweiten Forschung adressiert, indem an neuromorpher Hardware gearbeitet wird, die zunächst in der konventionellen Silizium-Halbleitertechnologie (CMOS) entwickelt wird. Für das breite Anwendungsspektrum des Digitalzeitalters haben diese Systeme langfristig jedoch immer noch einen viel zu hohen Energieverbrauch. Daher arbeiten Wissenschaftlerinnen und Wissenschaftler im Profilbereich ICT parallel an disruptiven Ansätzen für zukünftige Rechnerarchitekturen. Disruptiv deshalb, weil tief in die grundsätzliche Architektur der Rechner eingegriffen wird, sodass beispielsweise die synaptische Funktion nicht mehr

per Software nachempfunden, sondern direkt in einer neuartigen Hardware realisiert wird.

Memristive Bauelemente

Konkret werden neue Hardwarekomponenten für den Einsatz in künstlichen neuronalen Netzen erforscht. In den angestrebten Hardwarekomponenten sind die logischen Einheiten und die Speichereinheiten nicht mehr voneinander getrennt. Stattdessen wird im Speicher gerechnet, es wird ein Rechnen-im-Speicher realisiert. Dies ermöglicht eine massive Parallelisierung bei gleichzeitiger Vermeidung des energieintensiven und zeitaufwendigen Datentransfers über das von Neumann-Bottleneck. Spezielle memristive Bauelemente erfüllen diese Funktion. Memristor ist ein Kofferwort aus den englischen Wörtern „memory“ und „resistor“, die memristiven Bauteile haben also ein „Widerstandsgedächtnis“. Durch die Möglichkeit analoger Widerstandsänderungen stellen sie für das KI-Anwendungsfeld ideale Komponenten dar. Sie erfüllen zusätzlich weitere wichtige Anforderungen: platzsparende Geometrie, Kompatibilität zu Halbleiterfertigungsprozessen, Energieeffizienz und Skalierbarkeit. Auch ermöglichen sie eine effiziente Schnittstelle zu analogen Sensorsignalen, sodass die energieverbrauchende Analog-zu-digital-Umwandlung entfallen kann.

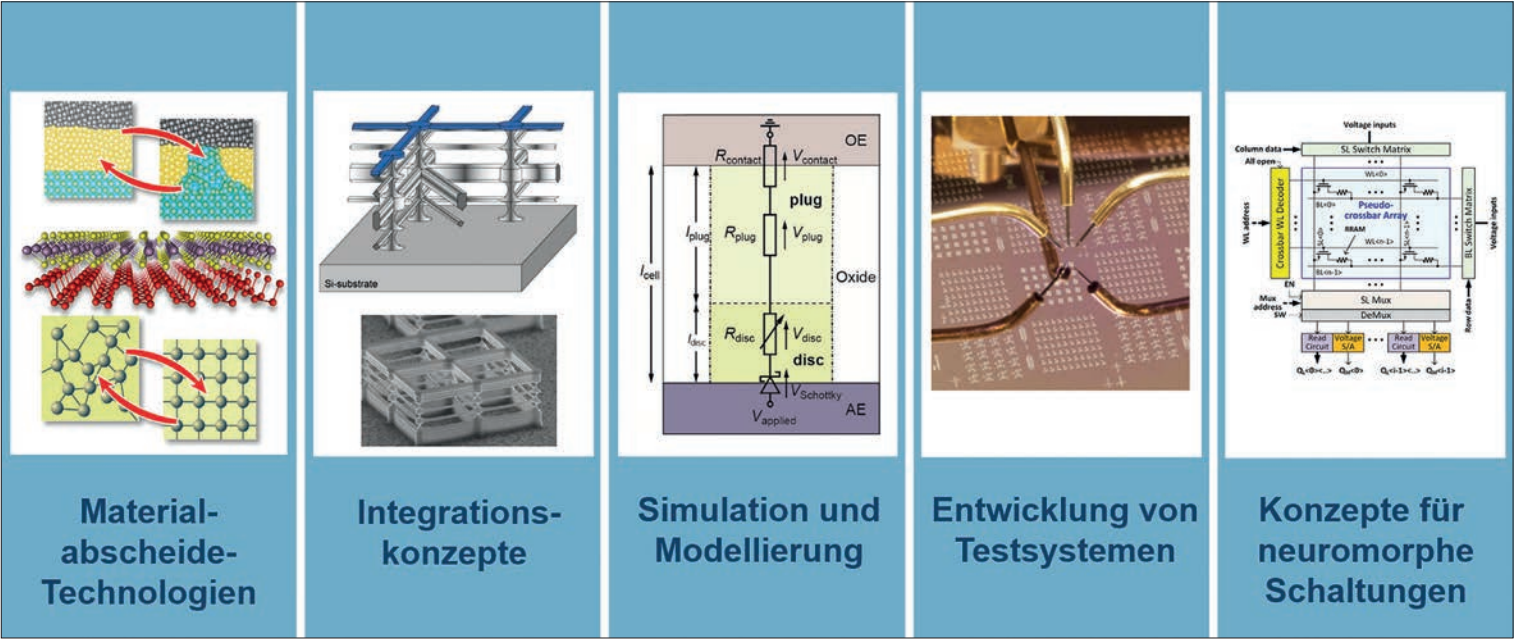


Bild 5: Arbeitsbereiche im Projekt „Neuroinspirierte Technologie der künstlichen Intelligenz für die Elektronik der Zukunft“, kurz NEUROTEC

Im Unterschied zu klassischen ladungsba-
sierten Speicherelementen wie dynamischen
Speichern oder Flash-Speichern basiert das
Konzept memristiver Bauelemente auf einer
Widerstandsänderung, die über eine atomare
Konfigurationsänderung der Speicherzelle
durch ein Spannungssignal hervorgerufen
wird. Abhängig vom physikalischen Mecha-
nismus und von den eingesetzten Materia-
lien können memristive Bauelemente als
nichtflüchtige oder semiflüchtige Speicher
ausgelegt werden, was sie für unterschiedliche
Einsatzbereiche in künstlichen neuronalen Net-
zen qualifiziert. Im Sonderforschungsbereich
„Neuroswitches“ werden – mit Förderung
durch die Deutsche Forschungsgemeinschaft
– die wichtigsten Schaltmechanismen mem-
ristiver Bauelemente, Phasenwechsel und Va-
lenzwechsel erforscht. Die gewonnenen Er-
kenntnisse zu den fundamentalen Prozessen
ermöglichen das gezielte Design memristiver
Bauelemente für spezifische Applikationen.
Bild 4 zeigt das mögliche Funktionsprinzip.
Die memristiven Elemente an den Kreuzungs-
punkten zweier Leiterbahnen können ähnlich
arbeiten wie die Synapsen in biologischen
Nervenzellen. Das Gewicht einer künstlichen
Synapse lässt sich durch Spannungspulse
wahlweise binär oder quasianalog einstellen.

Neuartige neuromorphe Hardware für autonome, selbstlernende Systeme

In der direkten Wechselwirkung mit der
physikalischen Umwelt ermöglichen neuro-
morph arbeitende Schaltungen eine hoch
effiziente Interaktion durch eine „natürliche“
von Ereignissen getriebene Verarbeitung. In
einem nächsten Schritt werden die kogniti-
ven Prozesse der Großhirnrinde in Hardware
abgebildet. Der genaue Ablauf komplexer
Prozesse im Gehirn ist selbst für Neurowis-
senschaftler noch ein Geheimnis. Sehr gut
verstanden sind allerdings die wesentlichen
Grundprinzipien in der „Signalverarbeitung“:
Stochastizität, Plastizität, Asynchronität, Kon-
nektivität und Heterogenität. Diese dienen
als Anregung für die Konzeption sogenann-
ter neuroinspirierter Systeme. Ein bereits
erwähntes Beispiel ist das Rechnen-im-Spei-
cher, bei dem Rechenoperationen im direkten
Verbund mit dem Speicherzellenfeld erfolgen.
Spannend wird es, wenn der beschriebene
Einsatz memristiver Bauelemente für neuar-
tige neuromorphe Computer mit der Ebene
der Algorithmen verbunden wird, in der
ebenfalls neuroinspirierte Prinzipien Anwen-
dung finden.
In der Biologie sind fehlerhafte Prozesse, zu-
fällige Ausfälle und stark verrauschte Daten
die Regel. Trotzdem, oder gerade deswegen,
beeindruckt die Natur immer wieder, zu

welchen Dingen sie fähig ist. Approximative
Berechnungen und ereignisgesteuerte Ver-
arbeitung dünnbesetzter Matrizen sind erste
Schritte, die signifikant bessere Energieeffi-
zienz auf moderne Rechnerarchitekturen zu
übertragen. Letztendlich sind Ingenieure
und Physiker aber noch gemeinsam mit den
Neurowissenschaftlern auf der Suche nach
dem „Berechnungsalgorithmus“ im mensch-
lichen Gehirn. Ein tiefgreifendes Verständnis
dieser kognitiven Prozesse ist Quelle für viele
Anregungen und Innovationen im Bereich
neuer Bauelemente, innovativer Schaltungs-
konzepte und revolutionärer Rechnerarchitek-
turen. Im Vergleich zu den beeindruckenden
Fortschritten im Bereich der künstlichen
neuronalen Netze sind die visionären Arbeiten
hin zu natürlicher Intelligenz noch auf einem
weiten Weg. Dennoch sind die erzielten
Forschungsergebnisse vielversprechend und
Innovationsmotor auf vielen Ebenen hin zu
den autonomen, selbstlernenden Systemen.

Vom Bauteil über Schaltkreise zur Anwendung

Die Expertise der beteiligten Institute ermög-
lichte die Einwerbung des Projekts „Neuro-
inspirierte Technologien der künstlichen In-
telligenz für die Elektronik der Zukunft“, kurz
NEUROTEC, beim Bundesministerium für
Bildung und Forschung. Hier werden seit



Bild 6: Beladung des Diffusionsofens mit Siliziumscheiben von 150 mm Durchmesser.

Foto: Peter Winandy

2019 schwerpunktmäßig Technologien zur Realisierung neuromorpher Rechnerarchitekturen adressiert. Zusammen mit regional ansässigen Firmen der Hochtechnologie, der aixACCT GmbH, der AIXTRON SE, der Surface systems + technology GmbH & Co. KG und der Synopsys GmbH, arbeiten die Wissenschaftlerinnen und Wissenschaftler von RWTH und Forschungszentrum Jülich an neuen memristiven Materialien, neuen Nanotechnologie-Prozessen, neuartigen Simulationswerkzeugen und Charakterisierungsmethoden und an Konzepten der Integration der memristiven Bauelemente in CMOS-Chips

für neuromorphe Rechner, siehe Bild 5. Die generischen neuromorphen Konzepte bauen aufeinander auf: Dies sind die erwähnten memristiv umgesetzten Rechnen-im-Speicher-Matrizen und assoziative Speicherblöcke sowie Methodiken zum Design neuromorpher Systeme basierend auf ebendieser Hardware. Weitere Aktivitäten finden sich in den kompetitiven FET-Open-Projekten QUEFORMAL und WIPLASH, in denen Wissenschaftler und Wissenschaftlerinnen von RWTH und AMO ebenfalls Aspekte des Rechnens-im-Speicher und der Plastizität erforschen. Im H2020-Projekt MNEMOSENE und im ITN-Projekt

MANIC arbeiten Wissenschaftlerinnen und Wissenschaftler von RWTH und Forschungszentrum Jülich zu Aspekten des neuromorphen Rechnens. Die Institute der RWTH und am Forschungszentrum Jülich sind nicht nur gut aufgestellt für die Forschung an disruptiven Technologien zur Realisierung von KI-Systemen, sondern auch in der Entwicklung neuartiger Algorithmen und der Erforschung der Anwendung von KI-Systemen. In den kommenden Jahren sollen diese beiden Bereiche „Technologien für neuromorphe Hardware“ und „Künstlich intelligente Systeme“ weiter ausge-



baut und noch stärker mit einander vernetzt werden. Auf der Anwenderseite ermöglichen neuromorphe Rechnerkonzepte die Entwicklung energieeffizienter oder energieautarker Systeme. Darunter fallen Objekterkennungssysteme, die in Verkehr und Logistik von zentraler Bedeutung sind sowie selbstlernende Systeme für den Einsatz in der Spracherkennung und in der Medizintechnik. Denkbare Anwendungen sind hier Cochlea-Implantate im Körper, Hörgeräte am Körper, tragbare Blutzuckerregulationssysteme oder Systeme zur Schmerzmittelapplikation.

Autoren

Univ.-Prof. Dr.-Ing. Tobias Gemmeke ist Inhaber des Lehrstuhls für Integrierte digitale Systeme und Schaltungsentwurf.

Dr.-Ing. Susanne Hoffmann-Eifert ist Gruppenleiterin am Institut für energie-effiziente Informationstechnologie des Forschungszentrums Jülich.

Dr. rer. nat Alexander Krüger ist Projektleiter des BMBF-Projektes NEUROTEC am Forschungszentrum Jülich.

Univ.-Prof. Dr.-Ing. Max Lemme ist Inhaber des Lehrstuhls für Elektronische Bauelemente und Geschäftsführer der AMO GmbH.

Univ.-Prof. Dr.-Ing. Rainer Waser ist Inhaber des Lehrstuhls für Werkstoffe der Elektrotechnik II und Leiter des Instituts für elektronische Materialien am Forschungszentrum Jülich.

Dr. ir. Dirk Wouters ist Gruppenleiter am Lehrstuhl für Werkstoffe der Elektrotechnik II.

Vom Hardware-Trojaner zum Blackout

Neue Cybersecurity-Methoden für Energienetze und Mikroprozessoren

Today, most computer users are aware that their laptops and smartphones may be attacked by cyber-criminals over the internet. Both the press and state authorities frequently issue warnings about newly detected vulnerabilities in hardware and software. Protective measures like virus scanners and regular software updates already prevent many individual attacks. Moreover, respecting simple security rules like carefully handling unknown email attachments or smartphone apps help to avoid that malicious software programs, such as Ransomware, acquire control over a user's machine. However, the necessary defense against cyber-attacks has many more dimensions than just protecting enduser devices. While attacks on those devices are annoying, yet cause limited economical damage, what if cyber-criminals took over energy infrastructures in order to blackmail or shut down an entire region or nation? What if the microprocessors underlying the hardware of all computers had inherent vulnerabilities that cannot be patched by updates in the field? This article looks at such novel threat scenarios and points out state-of-the-art counter-measures.

Die meisten Cyberangriffe basieren auf Sicherheitslücken in den Betriebssystemen wie Windows, Linux/Android oder iOS. Die Softwarekomplexität dieser Systeme ist extrem groß. Sicherlich gibt es noch etliche tausend unentdeckte Schwachstellen, die zu immer neuen Cyberattacken führen können. Ein Angreifer kann unbefugt die Kontrolle des Systems übernehmen, Daten ausspionieren, oder das System durch eine „Denial-of-Service“-Attacke lahmlegen. Erschwerend kommt hinzu, dass in letzter Zeit sogar Angriffe auf die Hardware der Computersysteme bekannt wurden, insbesondere auf die Prozessoren. Diese Central Processing Units, kurz CPUs, sind das Herzstück jedes PCs. Aus der Presse bekannte Beispiele sind „Spectre“ und „Meltdown“, welche Schwachstellen in Intel-CPU's ausnutzen und über verdeckte Kommunikationskanäle unbefugten Zugriff auf geschützte Speicherbereiche erlauben. Auch andere CPU-Hersteller wie ARM sind von ähnlichen Problemen betroffen, und somit potenziell Milliarden von Smartphones und andere Computer im Bereich Industrie 4.0 und dem Internet der Dinge.



Bild 1: Prüfung eines Edge Device für Stromnetze.
Foto: Peter Winandy

Wie bei den Betriebssystemen resultiert die Angreifbarkeit von CPUs in erster Linie aus deren Komplexität. Besteht die Hardware aus Milliarden von Transistoren, so sind unentdeckte Schwachstellen zahlreich vorhanden. Diese sind von den Herstellern in aller Regel nicht beabsichtigt. Allerdings ist es durchaus möglich, während des Chip-Entwurfs oder der Fertigung Angriffspunkte vorsätzlich in eine CPU einzuschleusen. Diese werden landläufig als „Hardware-Trojaner“ bezeichnet, sie bewirken eine unerwünschte und verdeckte Funktionalität in der Hardware, um Cyberangriffe zu unterstützen. Beispielsweise könnte eine CPU in einem Flugzeug durch eine bestimmte und im Normalbetrieb extrem selten auftretende Kombination von Instruktionen und Daten gezielt außer Betrieb gesetzt werden, um einen Denial-of-Service zu bewirken. In kritischen Infrastrukturen wäre ein solches Szenario offensichtlich fatal. Glücklicherweise sind Angriffe auf die Hardware wesentlich aufwendiger als die Ausnutzung von Software-Sicherheitslücken. Ein krimineller Durchschnittshacker verfügt nicht über das Wissen und die Mittel hierfür. Dies gilt jedoch nicht für staatliche Stellen, welche

aufgrund ihres Budgets Hardware-Trojaner realisieren könnten. Man denke hierbei nur an die weltweite Diskussion über das chinesische Unternehmen Huawei Technologies als möglichen Ausrüster für die kommenden 5G-Kommunikationsnetze. Auch im Bereich der Verteidigung ist die Diskussion über Hardware-Trojaner und entsprechende Abwehrtechniken in vollem Gange. So ist es nicht überraschend, dass das Thema Hardware-Security in den letzten Jahren in den Mittelpunkt geraten ist. Am Institut für Kommunikationstechnologien und eingebettete Systeme wurde im Rahmen eines Projektes erforscht, wie man eine typische CPU gegen Hardware-Trojaner und insbesondere Denial-of-Service-Angriffe absichern kann. Hierbei wurde angenommen, dass die Spezifikation und die frühen Entwurfsphasen in einer vertrauenswürdigen Umgebung stattfinden. Zu späteren Zeitpunkten kommen jedoch externe, möglicherweise global verteilte Partner für den detaillierten CPU-Entwurf sowie die Chip-Fertigung ins Spiel, welche als potenzielle Angreifer einzustufen sind. Um die Methodik der Absicherung mittels Open-Source-Lizenzbedingungen möglichst

transparent zu machen, fiel die Wahl auf den RISC-V, eine ursprünglich von der University of California in Berkeley stammende und sehr populäre neue CPU-Architektur. Darüber hinaus besitzt RISC-V eine geringe Komplexität, wodurch viele verdeckte Kommunikationskanäle von vornherein ausgeschlossen sind. Wie verhindert man jedoch, dass externe Partner und Dienstleister entlang der Lieferkette unbemerkt Hardware-Trojaner einschleusen können, obwohl sie zwangsläufig in den Besitz bestimmter Entwurfsdokumente kommen müssen? Der am Institut für Kommunikationstechnologien und eingebettete Systeme gewählte Ansatz basiert auf einer neuartigen Logikverschlüsselung. Diese verschleiert die Funktionsweise der CPU auf der Ebene der Logik-Gatter, sodass ein potenzieller Angreifer keine sinnvollen Ansatzpunkte für Hardware-Trojaner finden kann. Nur bei Kenntnis eines geheimen Schlüssels produziert die CPU sinnvolle Ergebnisse. Nach der Fertigung werden die Chips durch Programmierung der Schlüssel in vertrauenswürdiger Umgebung einmalig aktiviert. Zu diesem Zeitpunkt ist der Einbau neuer Hardware-Trojaner nicht mehr möglich, wodurch letztlich die ge-

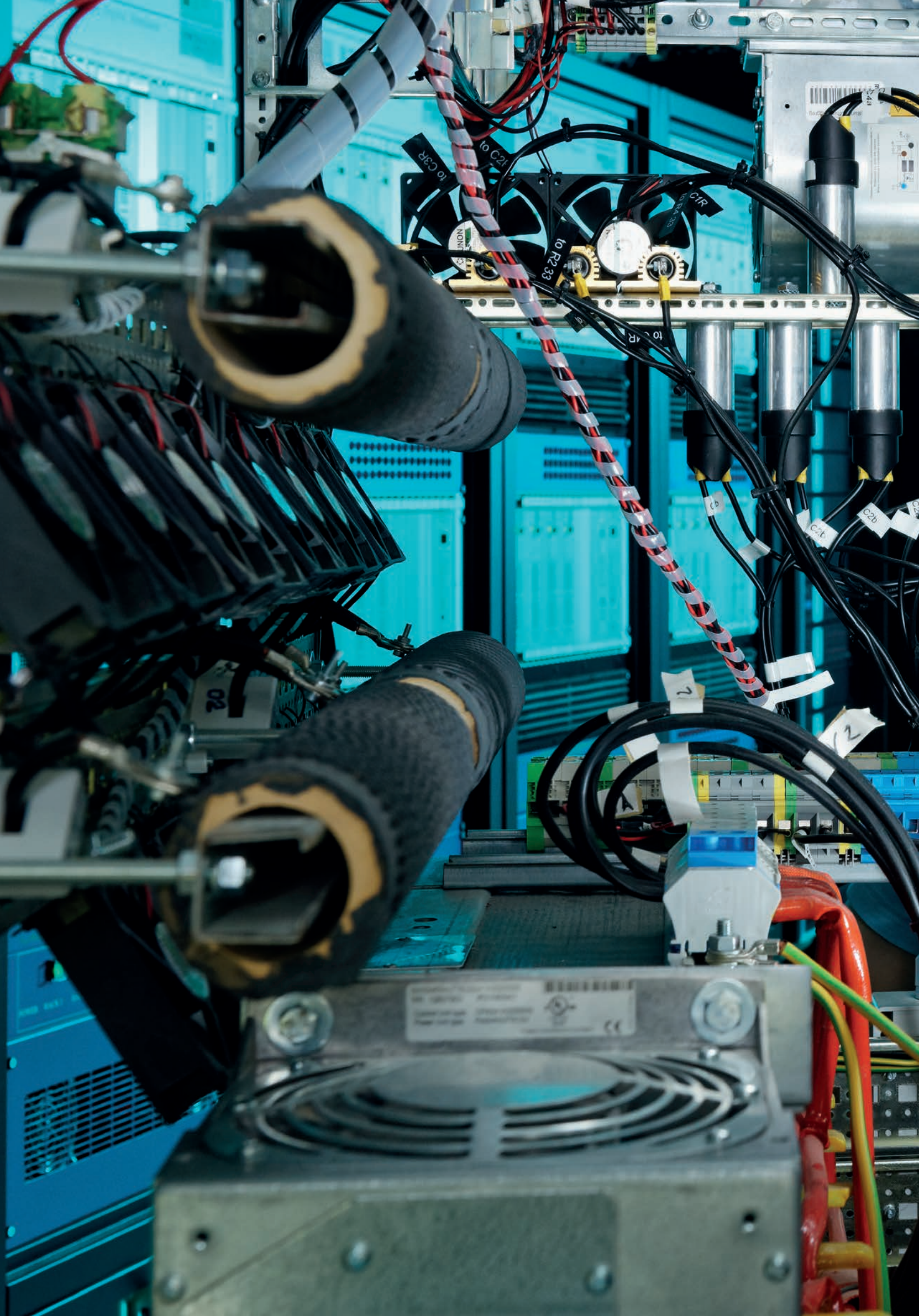




Bild 2: Stromnetz-Emulator am Lehrstuhl für Automation of Complex Power Systems
Foto: Peter Winandy

samte Lieferkette abgesichert ist. Bild 4 zeigt den Prototypen des abgesicherten RISC-V, welcher die Machbarkeit der Logikverschlüsselung bei sehr geringen Zusatzkosten in der Hardware nachweist. Da sich allerdings auch die technischen Möglichkeiten der Angreifer im Laufe der Zeit immer weiter verbessern, wird das Thema Hardwaresecurity weiterhin im Fokus der Forschung bleiben.

Durch die Energiewende wird der Anteil erneuerbarer Energien stetig erhöht. Dies bedeutet aber auch, dass Energiequellen eingesetzt werden, deren Beitrag zur Energieversorgung gegenüber konventionellen Energiequellen variiert. Daher ist eine automatisierte Bestimmung des Netzzustandes sowie prognostizierte Energiegewinnung und -verbrauch nötig, um eine stabile und nachhaltige Energieversorgung zu gewährleisten. Am Lehrstuhl für Automation of Complex Power Systems wird eine kostengünstige „Phasor Management Unit“, kurz PMU, entwickelt, welche den Netzzustand misst, die Ergebnisse in eine verteilte Serverstruktur sendet und dort weiterverarbeitet, um das Stromnetz zu regeln. Man spricht häufig von „Cloud Computing“, wenn Ressourcen mit anderen Projekten geteilt werden und nach Bedarf skalierbar sind.

Die Energieversorgung steht zunehmend im Fokus von Angriffen. Durch Manipulation der Infrastruktur sowie durch Fälschung der Messergebnisse kann die Regelung des Netzes nachhaltig gestört werden, was wiederum zu weitreichenden Problemen im öffentlichen Leben führen kann. Sowohl zur Berücksichtigung eines hohen Kostendrucks, wie er beispielsweise im Bereich der PMU vorliegt, als auch im Bereich des Cloud Computings

kommen häufig Standardbetriebssysteme zum Einsatz. Beim Cloud Computing werden die Anwendungen innerhalb eines Containers oder einer Virtuellen Maschine gestartet und teilweise mit einem vollständigen Betriebssystem ausgeliefert. Dies hat den Vorteil, dass der Dienstanbieter komplett von dem Infrastrukturanbieter getrennt werden kann. Allerdings sind die einzelnen Anwendungen innerhalb der virtualisierten Umgebungen sehr einfach gehalten, da hierdurch unter anderem eine Lastverteilung sehr einfach zu realisieren ist. Ein überlastetes System kann schnell durch einen zusätzlichen Knoten erweitert und die Anwendung ohne Störung des Dienstes dorthin migriert werden. Es stellt sich jedoch die Frage, ob innerhalb der virtualisierten Umgebung ein komplexes Mehrbenutzer- und Mehrprozess-Betriebssystem sinnvoll ist, wenn tatsächlich nur eine Anwendung innerhalb des Systems zum Einsatz kommt.

Wie bei den Prozessoren ist auch im Bereich der Betriebssysteme die Reduzierung der Komplexität und somit die Spezialisierung eine Vereinfachung der Verifizierbarkeit sowie eine Verkleinerung des Angriffsvektors. Die Spezialisierung kann so weit durchgeführt werden, bis ein komplettes Betriebssystem als Bibliothek zur Anwendung hinzugebunden wird und somit die Anwendung direkt auf einer Hardware gebootet werden kann. Solche Ansätze sind als „Unikernels“ bekannt und bieten durch die Spezialisierung eine höhere Leistungsfähigkeit gegenüber Standard-Betriebssystemen an. Zudem können Analysetechniken aus dem Bereich der Anwendungsentwicklung auf den kompletten Software-Stack – inklusive Kernel und Netzwerk-Stack – angewendet und somit Probleme besser erkannt werden.

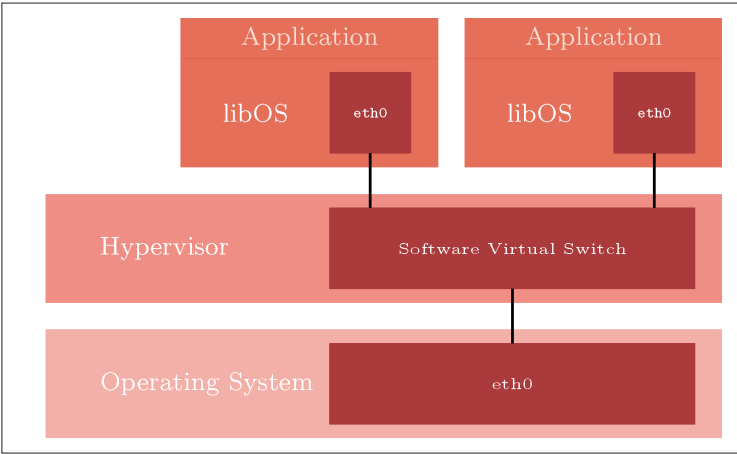
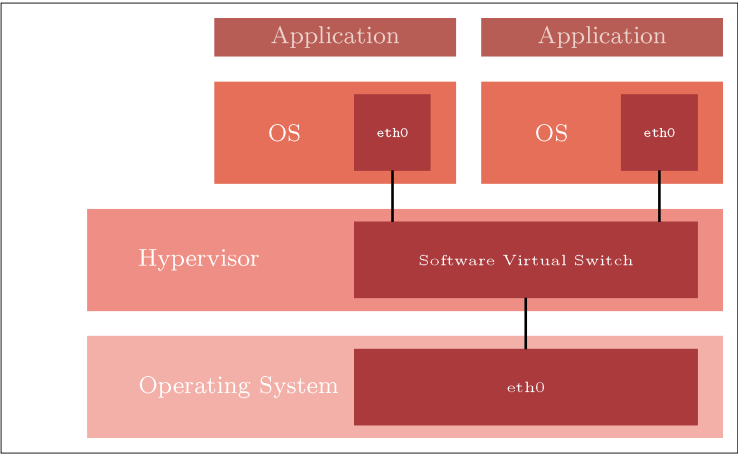


Bild 3: Klassischer Aufbau einer Cloud-Infrastruktur

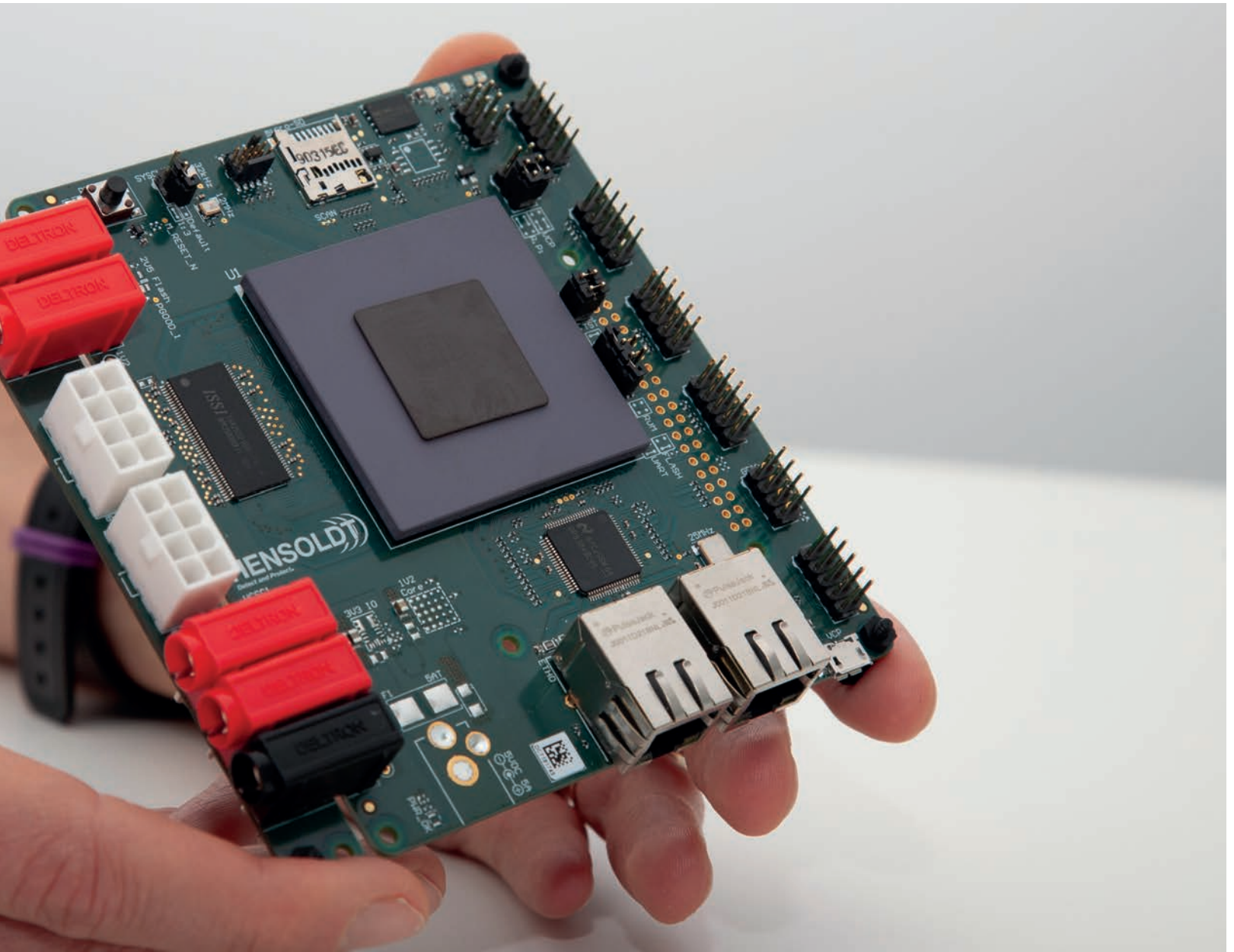


Bild 4: Sicherer „Made in Germany“ RISC-V Prozessor.
Quelle: Hensoldt Cyber GmbH

In der Betriebssystem-Entwicklung wird typischerweise C als Programmiersprache verwendet, die Dennis Ritchie vor rund 50 Jahren für die Entwicklung von Unix entworfen hat. Konzepte für einen sicheren und effizienten Umgang mit Arbeitsspeicher sind nicht vorhanden. Durch ihr schwaches Typsystem gibt es zudem zu viele Möglichkeiten, unsicheren Code zu schreiben. Daher wurde am Lehrstuhl für Automation of Complex Power Systems ein Unikernel entwickelt, der in der Programmiersprache Rust geschrieben wurde. Rust wurde im Rahmen der Entwicklung des Internet-Browsers „Fire-

fox“ entworfen und sollte sicher und schnell sein. Dies sind hervorragende Eigenschaften für die Entwicklungen eines Betriebssystems. Im EU-Projekt „Next Generation IoT as part of Next Generation Internet“ (IoT-NGIN), welches im Oktober 2020 startete, werden die entwickelten Konzepte für das Umfeld „Internet-of-Things“ ausgebaut, um so auch den Angriffsvektor für Sensoren und Kleinstgeräte zu reduzieren.

Autoren

Univ.-Prof. Dr. rer. nat. Rainer Leupers ist Inhaber des Lehrstuhls für Software für Systeme auf Silizium und ist im Leitungsgremium des Instituts für Kommunikationstechnologien und eingebettete Systeme. Univ.-Prof. Antonello Monti, Ph. D., ist Inhaber des Lehrstuhls für Automation of Complex Power Systems am E.ON Energy Research Center.



Bild 5: Schutz vor einer „Denial-of-Service“-Angriffe durch logikverschlüsselte Prozessoren.

Foto: Peter Winandy

ve been
ked

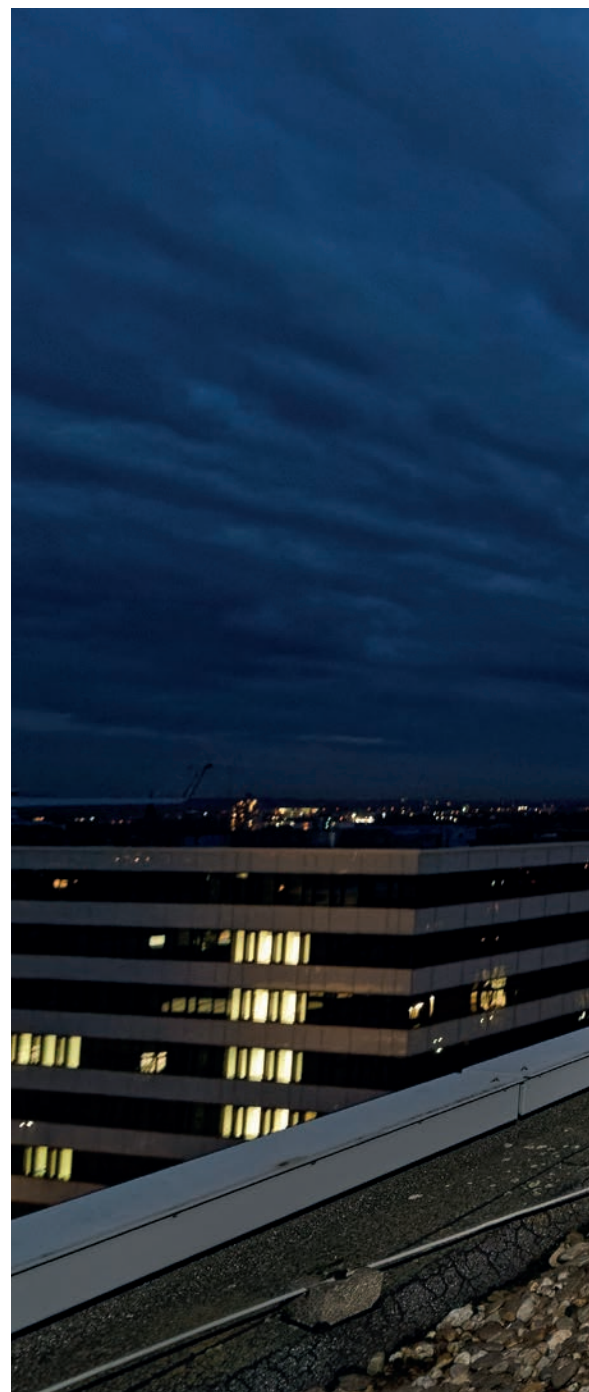


Von schwenkenden Beams und hohen Datenraten

Der 5G-Rollout aus Strahlenschutzsicht

The rollout of 5G in Germany has already started. To achieve high data rates even at the cell edge, massive MIMO antennas are used since they can radiate narrow beams to directions where service is needed. For safe operation, it must be guaranteed that the exposure of the general public to electromagnetic fields is compliant with legal safety limits in all publicly accessible areas and for all possible operational states. To monitor exposure, a measurement and extrapolation procedure is required to assess the maximum exposure when the base station is fully utilized. Such procedures exist for the established mobile radio networks GSM, UMTS and LTE. They are based on the exposure measurement of a reference signal and subsequent extrapolation to the maximum possible transmit power. In 5G, the Synchronization Signal Block (SSB) is an appropriate signal for extrapolation. However, in contrast to the previous mobile radio standards, the SSB can be radiated in different beams compared to the data traffic. They may vary in terms of gain, cell coverage, and orientations. A possible solution to determine the maximum exposure is the introduction of a location-dependent extrapolation factor, which takes into account the different characteristics of SSB and traffic beams at each measurement point. Measurements have shown that this approach leads to accurate results.

Das mobile Internet auf dem Smartphone gehört seit einigen Jahren zum Alltag vieler Menschen. Die Möglichkeit, Multimediainhalte auch unterwegs in hoher Qualität und Geschwindigkeit abrufen zu können, ist mehr und mehr zur Selbstverständlichkeit geworden. Dem dadurch stetig wachsenden Bedarf nach höheren Datenübertragungsraten begegnen die Mobilfunknetzbetreiber durch den Ausbau ihrer Netze und der Einführung neuer Mobilfunkstandards. Die Einhaltung der Immissionschutzgrenzwerte für hochfrequente elektromagnetische Felder hat hierbei eine entscheidende Bedeutung. Die Wissenschaftlerinnen und Wissenschaftler des Instituts für Hochfrequenztechnik forschen zur Erfassung und Abschätzung der Immission durch neue Mobilfunktechnologien. Im Zentrum steht die Verträglichkeit von Immissionen elektromagnetischer Felder mit der Umwelt und insbesondere Menschen. So wurden Untersuchungen an Small Cells vorgenommen, welche in Bereichen mit hohem Nutzeraufkommen für zusätzliche Datenübertragungskapazitäten sorgen. Auch wurden Immissionshöhen bei der Indoor-Versorgung mit Mobilfunk bestimmt, wie sie in Bürogebäuden und Flughäfen anzutreffen ist. Aktuell steht der Ausbau der neuen 5G-Technik im Mittelpunkt.



Der Mobilfunknutzer im Scheinwerferlicht

5G stellt in vielerlei Hinsicht neue Ansprüche an die Immissionsmessung. Eine der neuen technischen Möglichkeiten ist der Einsatz von Massive MIMO. MIMO steht für Multiple Input Multiple Output. Hiermit lassen sich Signale stark gebündelt in sogenannten Beams abstrahlen. Den Unterschied zur herkömmlichen Versorgung einer Funkzelle macht ein Vergleich deutlich: Eine Bühne kann als Ganzes durch einen breit strahlenden Scheinwerfer erhellt werden. Allerdings ist es auch möglich, einen schmalen helleren Lichtkegel auf eine



Bild 1: Durchführung einer Messung von Mobilfunkimmissionen an einem Dachstandort.
Foto: Peter Winandy

Person auf der Bühne zu richten und diese zu verfolgen. Genau das passiert auch bei Massive MIMO mit dem Funksignal: Es wird in Richtung eines zu versorgenden Endgeräts in einem Beam gebündelt und lässt sich schwenken. Auf diese Weise können Nutzer bis zum Rand der Funkzelle mit hohen Signalstärken versorgt und gleichzeitig die Störung anderer Nutzer reduziert werden. Dadurch lässt sich der Datendurchsatz erheblich steigern. Die Bestimmung der maximal möglichen Immission durch alle möglichen Beam-Konfigurationen stellt die Immissionsmesstechnik jedoch zusammen mit

weiteren Neuerungen in 5G vor erhebliche Herausforderungen. Im Rahmen eines durch das Bundesamt für Strahlenschutz geförderten Projekts wird an der Entwicklung eines geeigneten Verfahrens zur Erfassung der Maximalimmission unter diesen neuen Bedingungen geforscht.

Was zu beachten ist

Zur regelkonformen Inbetriebnahme einer Mobilfunkbasisstation sind die Grenzwerte der 26. Verordnung zur Durchführung des Bundes-Immissionsschutzgesetzes (26. BImSchV) einzuhalten. Die auftretenden Feldstärken

dürfen gemittelt über sechs Minuten auch bei höchster betrieblicher Anlagenauslastung nicht überschritten werden. Bei den bisherigen Mobilfunkgenerationen GSM, UMTS und LTE wird dieser Zustand erreicht, wenn alle Ressourcen zur Datenübertragung mit der höchsten vorgesehenen Sendeleistung genutzt werden. Durch die Einführung von Massive-MIMO-Antennen bei 5G kommt die Schwenkbarkeit der Beams hinzu. Die Gesamtleistung der Sendeanlage wird dabei auf die in der Funkzelle aktiven Endgeräte aufgeteilt. Maximalimmission wird deshalb nur erreicht, wenn alle Ressourcen mit maximaler

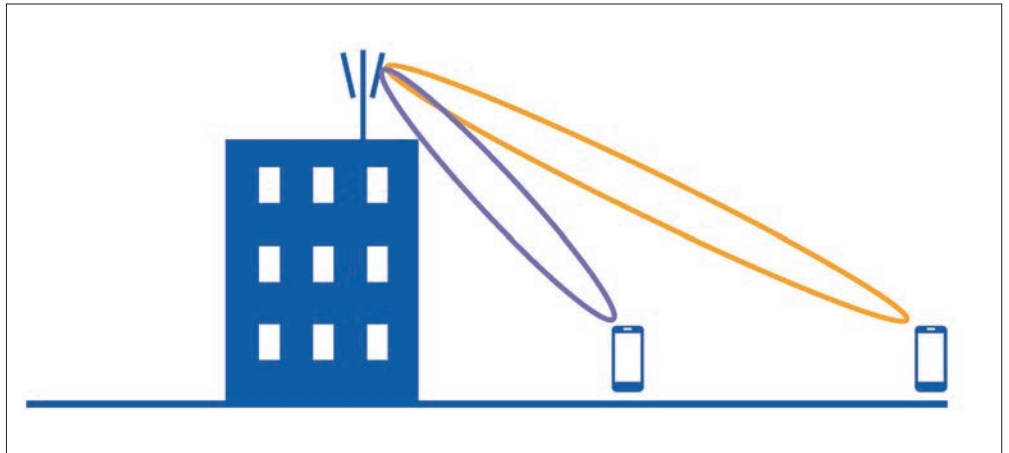


Bild 2: Mit der neuartigen Technologie 5G ist es möglich, einzelne Signale gebündelt in Beams innerhalb der Funkzelle zu schwenken und damit direkt auf einzelne Nutzer auszurichten. Dadurch kann eine hohe Signalstärke überall in der Zelle gewährleistet werden, ohne dass der Empfang anderer Nutzer gestört wird.

Sendeleistung in Form nur eines Beams abgestrahlt werden. Zusätzlich muss dieser für eine Dauer von mindestens sechs Minuten auf den Immissionsort einwirken. Dieser Betriebszustand ist sehr unwahrscheinlich, denn er erfordert, dass ein einzelnes Endgerät mindestens sechs Minuten lang einen Download mit einer Datenrate von einigen Gigabit pro Sekunde durchführt und gleichzeitig kein anderes Endgerät in der Zelle aktiv ist. Dennoch ist dieser Fall theoretisch möglich und muss für einen sicheren Betrieb der Basisstation berücksichtigt werden.

Wenn der Gewinn zum Problem wird

Der Zustand maximaler Anlagenauslastung kann also für eine Immissionsmessung in einer Mobilfunkzelle im Regelbetrieb nicht ohne Weiteres hergestellt werden. Daher muss die Maximalimmission auf anderem Wege bestimmt werden. Hierzu wird bei den bereits etablierten Mobilfunkstandards die Immission eines Signals gemessen, dessen Sendeleistung konstant ist, also nicht von der aktuellen Verkehrslast in der Zelle abhängt. Dazu eignen sich permanent ausgesandte Synchronisierungs- oder Referenzsignale. Sie dienen dem Endgerät zum Identifizieren der Funkzelle und zum Herstellen einer Ver-

bindung. Bei Kenntnis der Sendeleistung dieser Signale sowie der maximal möglichen Sendeleistung lässt sich eine Hochrechnung auf Maximalimmission durchführen.

Bei 5G existiert mit dem Synchronization Signal Block, kurz SSB, ein vergleichbares Signal, das immer wiederkehrend auch ohne Verkehrslast mit konstanter Sendeleistung abgestrahlt wird. Somit eignet es sich prinzipiell als Ausgangspunkt für eine Hochrechnung auf Maximalimmission. Die große Herausforderung besteht allerdings darin, die Strahlschwenkung adäquat zu berücksichtigen. Der SSB muss im Gegensatz zu den Vorgängerstandards nicht mehr zellweit abgestrahlt werden, sondern kann, wie Bild 3 zeigt, als Teil einer Folge bestehend aus bis zu acht aufeinanderfolgenden SSBs in verschiedene Richtungen ausgesendet werden. Dies dient zur Vorauswahl der Abstrahlrichtung für die nachfolgende Datenübertragung, denn die SSB Beams werden abhängig vom Aufenthaltsort des Nutzers innerhalb der Funkzelle unterschiedlich stark empfangen. Die eigentlichen Daten werden mithilfe von Traffic Beams ausgesandt. Diese können stärker gebündelt sein und im Gegensatz zum SSB mit einer deutlich feineren Auflösung in verschiedene Richtungen ausgerich-

tet werden. Nur im Falle eines einzelnen voll ausgelasteten Traffic Beams tritt Maximalimmission auf. Diese kann durch Berücksichtigung des Unterschieds zwischen der Bündelung des SSB und Traffic Beams berechnet werden. Ein Maß für die Bündelung ist der Antennengewinn, der umso höher ausfällt, je stärker ein Beam konzentriert ist. Nicht nur der Gewinn von SSB und Traffic Beams unterscheidet sich, sondern auch die Abdeckungsbereiche können, wie in Bild 4 angedeutet, voneinander abweichen. Dies hat zur Konsequenz, dass eine pauschale Hochrechnung auf Maximalimmission, die lediglich den Anteil der Sendeleistung vom SSB an der Gesamtsendeleistung berücksichtigt, nicht anwendbar ist.

Eine mögliche Lösung und deren Validierung

Ein vielversprechender Ansatz zur Berücksichtigung der unterschiedlichen Antennengewinne von SSB und Traffic Beams ist die Einführung eines ortsabhängigen Korrekturfaktors. Mithilfe der einfach zu bestimmenen relativen Horizontal- und Vertikalwinkel zwischen dem jeweiligen Messpunkt und der Basisstationsantenne lässt sich der Gewinnunterschied geeignet ermitteln. Dieses Ver-

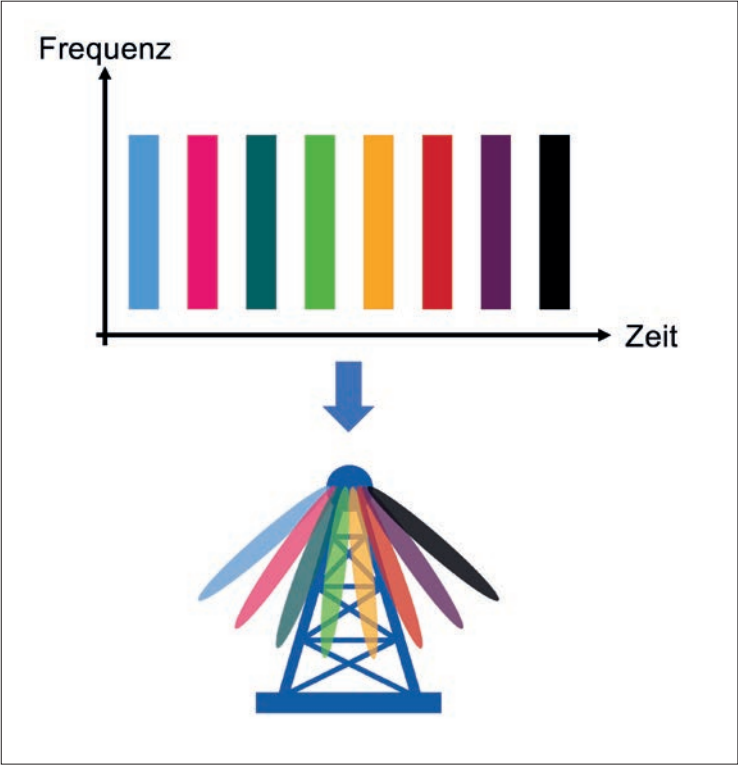


Bild 3: Der Synchronisation Signal Block (SSB) wird nacheinander als Folge aus bis zu acht (unterscheidbaren) SSB Beams in verschiedene Richtungen ausgesendet. Dies dient zur Vorauswahl der Abstrahlrichtung für die nachfolgende Datenübertragung, denn die SSB Beams werden abhängig vom Aufenthaltsort des Nutzers innerhalb der Funkzelle unterschiedlich stark empfangen.

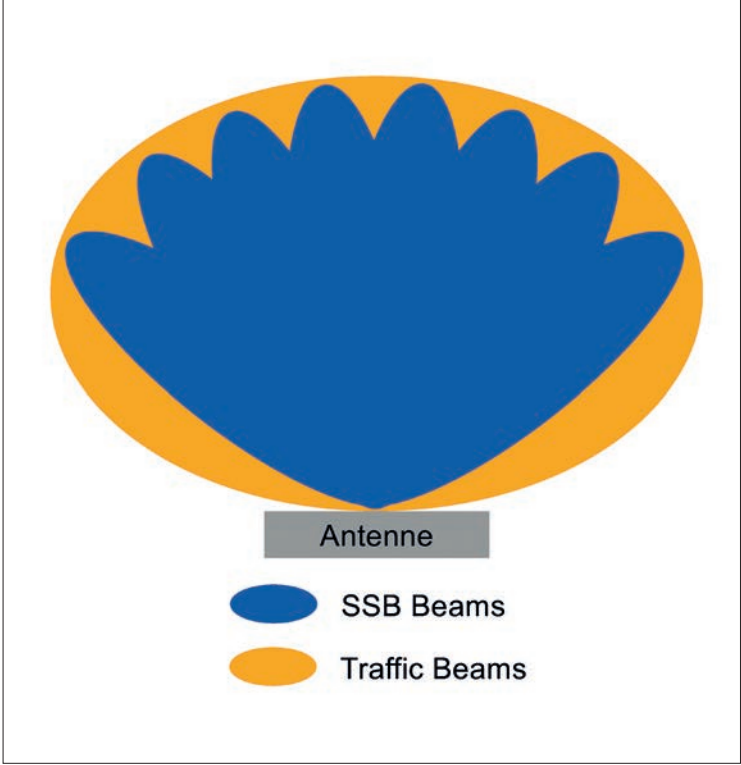


Bild 4: Die Abdeckungsbereiche der SSB Beams und der Einhüllenden aller möglichen Traffic Beams können sich voneinander unterscheiden. Deshalb kann nicht unmittelbar von der Messung eines SSB Beams auf die maximale Feldstärke des Traffics geschlossen werden.

fahren erfordert jedoch präzise Kenntnisse über das räumliche Abstrahlverhalten von SSB und Traffic Beams. Zur Validierung der auf diese Weise hochgerechneten Immission wurde zum Vergleich die Immission während tatsächlicher Volllastung der Basisstation gemessen. Hierfür wurde mit einem 5G-fähigen Endgerät ein einzelner vollausgelasteter Traffic Beam erzeugt. Es konnte belegt werden, dass die Hochrechnung mittels ortsabhängigem Korrekturfaktor zu Ergebnissen mit hoher Genauigkeit führt. Das bislang entwickelte Hochrechnungsverfahren zur Bestimmung der Maximalimmission erfordert umfangreiche Informationen, um einen ortsabhängigen Korrekturfaktor ableiten zu können. Zusätzlich können sich von Basisstation zu Basisstation die SSB und Traffic Beams hinsichtlich ihrer Anzahl, ihrer Abdeckungsbereiche sowie ihrer Antennengewinne unterscheiden. Langfristig ist daher ein praktikableres Hochrechnungsverfahren erstrebenswert, das auch ohne Kenntnis der jeweiligen Antennenkonfiguration eine verlässliche Aussage über die maximal mögliche Immission treffen kann. An einem entsprechenden generischen Hochrechnungsverfahren wird derzeit geforscht. Es soll ohne individuelle Korrektur der Antennengewinne

von SSB und Traffic Beams auskommen, sondern in bestimmten Bereichen der Zelle anwendbare pauschale Korrekturfaktoren erlauben. Dabei ist sicherzustellen, dass die Maximalimmission unter keinen Umständen unterschätzt, aber auch nur in akzeptablem Umfang überschätzt wird. Weitere Forschungsarbeiten streben die Umsetzung eines zeitkontinuierlichen und flächendeckenden Immissions-Monitorings an. In einer Machbarkeitsstudie wird ein Crowdsourcing-basierter Ansatz verfolgt, welcher die von Smartphones erfassten Signalstärkewerte zur Immissionsabschätzung heranzieht. Auch in einem weiteren Projekt mit Bezug zu 5G dürfen die Wissenschaftlerinnen und Wissenschaftler des Instituts für Hochfrequenztechnik ihre Expertise einbringen: Im Förderrahmen 5G.NRW wird im Vorhaben „Giga for Health: 5G-Uni-Medizincampus. NRW“ die Störfestigkeit von Medizingeräten bei der Ausstattung von Krankenhäusern mit 5G-Versorgung im Innenbereich untersucht.

Autoren

Univ.-Prof. Dr.-Ing. Dirk Heberling ist Inhaber des Lehrstuhls für Hochfrequenztechnik und Leiter des Instituts für Hochfrequenztechnik sowie Leiter des Fraunhofer-Instituts für Hochfrequenzphysik und Radartechnik. Thomas Kopacz, M. Sc., Dipl.-Ing. Sascha Schießl und Anna-Malin Schiffarth, M. Sc., sind wissenschaftliche Mitarbeiter am Institut für Hochfrequenztechnik.

Quanten- computing

Ein Paradigmenwechsel in der Datenverarbeitung

Quantum technologies have already found their way into many parts of our daily lives. Current developments in quantum computing are promising to bring us into a new era of computing power beyond our present imagination. RWTH Aachen University and its strong collaboration partners are on the way to making significant contributions to this expected paradigm shift in the future of computing. Benefiting from a number of European and national funding programs as well as collaborations with non-university research institutions, research groups at RWTH and JARA-FIT are expanding their competencies with the goal of contributing substantially to European and national R&D programs in the field. Research competencies span a broad portfolio in the theory of the elementary building blocks of quantum computers, quantum error correction, the experimental realizations of solid state approaches based on semiconductors and superconducting devices, as well as system engineering aspects such as cryoelectronics.

Quantentechnologien der ersten Generation sind längst Teil unseres Alltags geworden. Halbleiterbauelemente wie Dioden und Transistoren bilden die Grundlage der modernen Elektronik und sind in Computern, Smartphones und Fernsehern enthalten. Ohne Laser wären viele Fertigungsprozesse, Messverfahren und Kommunikation undenkbar. Auch viele materialwissenschaftliche Errungenschaften wie zum Beispiel Supraleitung basieren auf dem Verständnis der Quantenmechanik. All diese Anwendungen wären ohne letztere nicht denkbar. Jedoch sind die benutzten Phänomene stets makroskopischer Natur und beruhen auf dem kollektiven Verhalten einer großen Zahl von mikroskopischen Teilchen. Die Manipulation einzelner Quantenobjekte wie Atome, Elektronen oder auch Photonen erschließt weitere Quantenphänomene, welche anfänglich eher als praktisch belanglose Kuriosität erschienen, mittlerweile aber experimentell eindeutig untersucht und bestätigt sind. Deren Ausnutzung liegt im Fokus der Quantentechnologien der zweiten Generation. Für diese Quantenobjekte gelten physikalische Gesetze und Prinzipien, die das Potenzial für neue technische Lösungen bieten. Ein vielversprechender Entwicklungszweig der sogenannten „Quantum 2.0 Ära“ ist das

Quantencomputing. Aufgrund ihres bahnbrechenden Potenzials gelten Quantencomputer als eine Schlüsseltechnologie des 21. Jahrhunderts. Die konzeptionelle Grundlage beruht im Wesentlichen auf zwei Phänomenen: Überlagerungszustände und Verschränkung. Quantenbits, auch Qubits genannt, können wie das klassische Bit den Zustand 1 oder 0 annehmen – aber auch gleichzeitig im Zustand 1 und 0 sein sowie in beliebigen Zuständen dazwischen. Verfügt man über mehrere Qubits, können wiederum solche Superpositionen abgebildet werden, die jeden möglichen klassischen Zustand des Gesamtsystems enthalten. So kann ein einzelner Quantenprozessor Rechenschritte auf vielen Zuständen simultan erledigen. Da die Anzahl der möglichen Zustände exponentiell mit der Anzahl der Qubits steigt, gilt dies auch für die Rechenleistung, welche weit jenseits der klassischer Computer liegen kann. Wissenschaftlerinnen und Wissenschaftler erhoffen sich von Quantencomputern Durchbrüche in vielen rechenintensiven Disziplinen, etwa bei der Erforschung neuer Medikamente, bei der Entwicklung Künstlicher Intelligenz oder auch für die Optimierung von Lieferketten in der Logistik. Weiterhin ermöglicht die Eigenheit der Quantenmechanik, dass Zustände nicht ohne

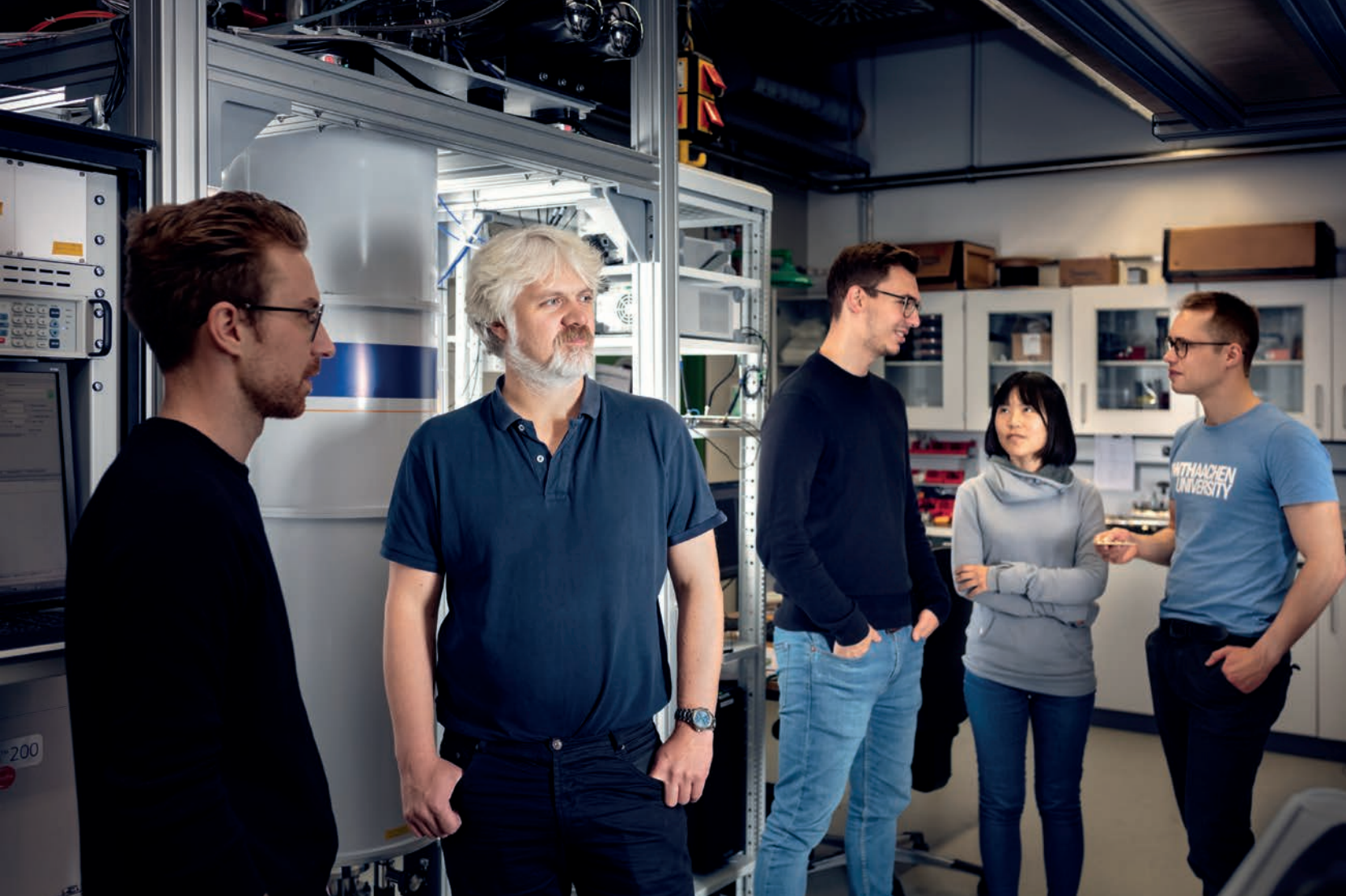


Bild 1: Im Labor widmen sich Nachwuchswissenschaftler um Professor Hendrik Blum (zweiter von links) der experimentellen Untersuchung von spinbasierten Qubits.
Foto: Simon Wegener

Veränderung kopiert oder gelesen werden können. Das Ergebnis sind Kommunikationsprotokolle, die physikalisch gegen unbemerktes Abhören geschützt sind.

Zur Quantenüberlegenheit ist es noch ein weiter Weg

Nach nahezu zwei Jahrzehnten Forschung an den Eigenschaften einzelner Qubits stehen Quantencomputer aktuell noch am Anfang ihrer Entwicklung. Bisher ist es gelungen, Rechensysteme mit bis zu 50 verschränkten Qubits zu realisieren. Dies ist für praktische Anwendungen jedoch noch deutlich zu wenig, und die Qubits arbeiten noch zu ungenau. Je nach Ansatz werden Verbesserungen von mehreren Größenordnungen in der Fehlerrate oder Zahl der Qubits benötigt. Die Schwierigkeit liegt in der sehr hohen Anfälligkeit von Quantenzuständen gegen Störungen jeder Art, so dass sie nur für kurze Zeit miteinander verschränkt bleiben und in dieser Zeit Rechenoperationen ausführen können. Man spricht von Dekohärenz, die die Fehlerrate bestimmt. Ein Ansatz, von dem relativ frühe Erfolge zu erhoffen sind, ist die Verwendung von bis zu einigen hundert Qubits für kurze Algorithmen, so dass Fehler selten genug auftreten. Dadurch kann das Potenzial jedoch nicht voll

ausgeschöpft werden. Für längere Algorithmen wird die sogenannte Quantenfehlerkorrektur notwendig, welche die unvermeidbaren Fehler korrigiert. Der Preis dafür ist eine Vergrößerung der notwendigen Qubit-Zahl um einen Faktor 1.000 bis 10.000. Noch keiner der verfolgten Implementierungsansätze für Qubits hat die Voraussetzungen für eine solche Hochskalierung bewiesen. Zu der Schwierigkeit der Realisierung hinreichend vieler robuster Qubits kommt die Realisierung geeigneter Kontrollsysteme. Jedes Qubit benötigt sorgfältig abgestimmte, hochpräzise Kontrollsignale. Für kleine Systeme bewährte Verfahren lassen sich nicht ohne weiteres hochskalieren. Manche Qubits werden bei Raumtemperatur im Ultrahochvakuum betrieben und erfordern genaue Laserpulse, welche sich nur mit sehr aufwendigen, makroskopischen Lasersystemen erzeugen lassen. Andere sind zwar mit einfacher zu erzeugenden elektrischen Signalen kontrollierbar, müssen aber nahe dem absoluten Nullpunkt bei -273 °C betrieben werden. Gängige Temperaturen sind ein Faktor 1.000 bis 10.000 kleiner als Raumtemperatur. Somit wird noch viel Grundlagenforschung und Entwicklungsarbeit zu einer Reihe von Aspekten benötigt, bevor verschiedene Anwendungsfelder des Quantencomputings

erschlossen werden können. Wenn dies gelingt, winkt jedoch eine großzügige Belohnung: Eine Studie der Boston Consulting Group schätzt ein langfristiges Marktvolumen von bis zu 300 Milliarden Dollar, wobei die Zuverlässigkeit dieser Vorhersage nicht höher liegen dürfte als die derzeitiger Qubits.

Globale und nationale Trends

Mit diversen nationalen Förderprogrammen investieren viele Regierungen, vor allem China und die USA, in Quantentechnologien. 2018 startete das FET Flagship Program on Quantum Technologies der Europäischen Kommission, um die Entwicklung der Quantentechnologien auch in Europa mit einer Milliarde Euro Fördergeldern voranzutreiben. Zusätzlich fördert die Bundesregierung die Entwicklung der Quantentechnologien in Deutschland mit 650 Millionen Euro bis 2021. Auch beschloss die Bundesregierung am 3. Juni 2020 ein Konjunkturprogramm, in welchem Quantentechnologien, insbesondere Quantencomputing, Quantenkommunikation, Quantensensorik und Quantenkryptographie, mit einem Finanzvolumen in Höhe von 2 Milliarden Euro gefördert werden sollen. Ziel ist, dass deutsche Institute und Unternehmen die zweite Quantenrevolution mitgestalten und international eine führende Rolle übernehmen. Große Konzerne wie Google, IBM und Microsoft haben bereits experimentelle Quantencomputer vorgestellt. Partnerschaften zwischen diesen Konzernen und den deutschen außeruniversitären Forschungseinrichtungen sind bereits entstanden: Während IBM seinen Quantencomputer „Q System One“ in einer Partnerschaft mit der Fraunhofer-Gesellschaft nach Deutschland bringt, wurde ein gemeinsames Forschungsprogramm zwischen dem Forschungszentrum Jülich und Google initiiert. Die „Jülich UNified Infrastructure for Quantum Computing“, kurz JUNIQ, wird für die Ausbildung von Spezialisten in der Industrie verfügbar und über die Cloud europäischen Nutzern zugänglich sein. Wissenschaftlerinnen und Wissenschaftler beider Partner bekommen zusätzlich die Möglichkeit, Simulationen auf den Superrechnern am Jülich Supercomputing Centre durchzuführen und mit Googles Quantenprozessoren zu experimentieren.

Die RWTH und ihre Partner haben in den letzten zehn Jahren den Forschungsschwerpunkt Quantencomputing systematisch aufgebaut. Ein wichtiges Element ist die Jülich Aachen Research Alliance, kurz JARA, die die Kompetenzen beider Einrichtungen

bündelt. JARA-FIT, eine der fünf Forschungssektionen, wurde den Zukunftstechnologien in der Datenverarbeitung gewidmet. Das JARA-FIT angehörige JARA Institut Quantum Information (QI) hat es sich gemeinsam mit Partnern zur Aufgabe gemacht, zum dynamischen Forschungsfeld der Quanteninformation maßgeblich beizutragen und auch in Europa Technologiekompetenz in Quantencomputing-Hardware aufzubauen. Die Professoren David DiVincenzo und Hendrik Bluhm leiten das Institut.

Die wesentlichen Forschungsschwerpunkte des sowohl theoretisch als auch experimentell arbeitenden Instituts sind das Verständnis von Dekohärenzeffekten in Festkörper-Qubits, die Theorie der Quantenfehlerkorrektur und die experimentelle Weiterentwicklung von supraleitenden und Halbleiterqubits mit zunehmendem Fokus auf Skalierbarkeit. So entsteht in Jülich der Systemdemonstrator des Projekts „OpenSuperQ“, das im Rahmen FET Flagship Program on Quantum Technologies der EU einen Quantenprozessor mit 50 bis 100 Qubits entwickelt. Im Bereich Halbleiterqubits wurde in Zusammenarbeit mit Infineon kürzlich ein konkreter Skalierungsansatz konzipiert, der mit industrieller Fertigung kompatibel und zum Erreichen sehr großer Qubit-Zahlen geeignet ist. Dieser wurde 2019 mit dem RWTH Innovation Award ausgezeichnet.

Ein relativ neuer Trend in der lokalen Forschungslandschaft ist die Erforschung der Grundlagen für hochintegrierte, kryoelektronische Kontrollsysteme, welche den Schlüssel zur Hochskalierung liefern könnten. Wichtige Forschungsfragen sind: Wie gut funktioniert die heutige Schaltungstechnik bei tiefen Temperaturen? Wie kann sie für die geänderten Anforderungen optimiert werden? Welche Schaltungs- und Systemkonzepte eignen sich für die skalierbare Qubit-Kontrolle?

Autoren

Univ.-Prof. Dr. rer. nat. Hendrik Bluhm ist Inhaber des Lehrstuhls für Experimentalphysik und im Leitungsgremium des JARA Institut Quantum Information.

Dr. Marian Barsoum ist zuständig für die Öffentlichkeitsarbeit im Exzellenzcluster „ML4Q – Matter and Light for Quantum Computing“ und am II. Physikalischen Institut der Universität zu Köln.



Namen & Nachrichten

DAAD-Preis 2020 für Wafic El Sabbagh

Rektor Ulrich Rüdiger zeichnete Wafic El Sabbagh mit dem DAAD-Preis 2020 aus. Der Deutsche Akademische Austauschdienst, kurz DAAD, würdigt mit der Verleihung hervorragende Leistungen internationaler Studierender an deutschen Hochschulen. Der Preis ist mit 1.000 EUR dotiert. El Sabbagh kommt aus dem Libanon und studiert seit dem Wintersemester 2018/19 im Bachelorstudiengang Bauingenieurwesen. Der 21-jährige überzeugt nicht nur mit seinen Prüfungsleistungen, sondern engagiert sich zusätzlich ehrenamtlich.

Erster grenzüberschreitender vollautomatischer Rettungsflug

Das Bundesministerium für Verkehr und digitale Infrastruktur finanziert seit 2019 das Forschungsprojekt „GrenzFlug“. RWTH und Stadt Aachen kooperieren hier zu Such- und Rettungsmissionen mit unbemannten Flugsystemen. Im November ist der erste grenzüberschreitende vollautomatische Rettungsflug gestartet. Automatisierte Flugsysteme können ein Gebiet schnell erreichen und auch schwer passierbare Gebiete systematisch absuchen. In Aachen wurden die Voraussetzungen für die Integration von unbemannten Luftfahrzeugen in grenzüberschreitende Rettungsmissionen erarbeitet. Eine Herausforderung lag in den unterschiedlichen luftrechtlichen Rahmenbedingungen im Dreiländereck. Am Institut für Flugsystemdynamik unter Leitung von Professor Dieter Moormann wurde ein unbemanntes Flugsystem mit den notwendigen Sensoren ausgestattet. Das Flugsystem kann als Kippflügler senkrecht starten, landen und gleichzeitig vollständig automatisiert große Strecken auch außerhalb der Sichtweite eines Steuerers zurücklegen. Eine Kamera überträgt auch Daten im Infrarotspektrum und kann so eine Vermisstensuche selbst bei Nacht und Nebel durchführen.

Bei der Demonstration hat das Flugsystem mehrfach die Grenzen zwischen den Niederlanden und Deutschland überflogen und damit immer wieder seine luftrechtliche Genehmigung geändert.

Kooperation mit University of Cape Town

Das Bundesministerium für Bildung und Forschung fördert das Projekt „VaReeWa: Entwicklung eines kombinierten pyro-/hydro-metallurgischen Prozesses zur Wertrückgewinnung aus Elektronikschrott durch kleinformatige Umsetzung in Afrika“. Im Rahmen des Deutsch-Afrikanischen Innovationsförderpreises wird die Zusammenarbeit von Professor Bernd Friedrich vom Lehrstuhl für Metallurgische Prozesstechnik und Metallrecycling und Professor Jochen Petersen von der University of Cape Town in Südafrika für 24 Monate unterstützt. Mit der Preisvergabe soll die Umsetzung von Forschungsergebnissen in afrikanischen Ländern ermöglicht sowie die Kooperation zwischen deutschen Wissenschaftlerinnen und Wissenschaftlern mit afrikanischen Partnern gestärkt werden.

Michael Schaub ist Tenure-Track-Juniorprofessor

Michael Schaub baut als Tenure-Track-Juniorprofessor in der Fachgruppe Informatik eine Nachwuchsgruppe im Bereich „Computational Network Science“ auf. Er überzeugte mit seinem Antrag „Learning from networks with unobserved edges“ im Auswahlverfahren des „Programms zur Förderung der Rückkehr des hochqualifizierten Forschungsnachwuchses aus dem Ausland“ des Landes Nordrhein-Westfalen und wird die nächsten fünf Jahre mit 1,25 Millionen Euro unterstützt. Schaub studierte Elektrotechnik und Informationstechnologie an der ETH Zürich und promovierte 2014 am Imperial College London

in Mathematik. Anschließend folgten Forschungstätigkeiten an der Université catholique de Louvain und der Université de Namur in Belgien, dem Massachusetts Institute of Technology und der University of Oxford. An der RWTH arbeitet Schaub an der Analyse komplexer Systeme und Netzwerke.

Frauen-MINT-Award 2020 für Ina Berenice Fink

Ina Berenice Fink vom Lehrstuhl für Informatik 4 (Kommunikation und verteilte Systeme) wurde mit dem Frauen-MINT-Award 2020 ausgezeichnet. Die audimax Medien GmbH und die Deutsche Telekom AG würdigen mit diesem Preis die besten Bachelor- und Masterarbeiten weltweit von Absolventinnen der MINT-Studiengänge. MINT steht für Mathematik, Informatik, Naturwissenschaften und Technik. In der Wettbewerbsrunde „Wachstumsfelder der Zukunft“ lag der Fokus auf den Themen Cloud, Internet der Dinge, Künstliche Intelligenz, Cyber Security und Netze der Zukunft. Neun Preise wurden vergeben, Fink erhielt einen der beiden Gesamtpreise. Ihre Arbeit „Advanced In-Network Security for Smart Devices“ ist im Bereich Cyber Security angesiedelt und entstand am Lehrstuhl für Informatik 4 (Kommunikation und verteilte Systeme) unter Leitung von Professor Klaus Wehrle.

Forschungspreis für Walter Leitner

Professor Walter Leitner, Inhaber des Lehrstuhls für Technische Chemie und Petrolchemie, erhält den Georg Wittig-Victor Grignard-Preis. Dieser wird jährlich an exzellente Chemikerinnen und Chemiker verliehen, die die wissenschaftliche Vernetzung zwischen Deutschland und Frankreich fördern. Die Société Chimique de France und die Gesellschaft Deutscher Chemiker wählen im Wechsel die Preisträgerin beziehungsweise den Preisträger. Die Société Chimique de France nominierte Leitner. Gewürdigt werden seine Forschungen zur Homogenen Katalyse in überkritischem Kohlenstoffdioxid und zur chemischen Umwandlung und Nutzung von Biomasse. Leitner ist seit 2002 Inhaber des Lehrstuhls für Technische Chemie und Petrolchemie, seit Oktober 2017 ist er auch Direktor am Max-Planck-Institut für Chemische Energiekonversion in Mülheim an der Ruhr. Gemeinsam mit seinem Team erforscht er katalytische Prozesse von den molekularen Grundlagen bis zu reaktionstechnischen Konzepten. Für seine Beiträge auf diesem Gebiet war er zusammen mit Industriekollegen der

Firma Covestro auch 2019 unter den Finalisten für den Deutschen Zukunftspreis.

Fernsteuerung von Binnenschiffen

Im Institut für Regelungstechnik unter Leitung von Professor Dirk Abel werden im Verbundprojekt „FernBin“ Methoden und Algorithmen entwickelt, um Binnenschiffe sicher und effizient fernzusteuern. Das Bundesministerium für Wirtschaft und Energie fördert das Projekt, Partner sind das Entwicklungszentrum für Schiffstechnik und Transportsysteme e.V. in Duisburg, die Universität Duisburg-Essen, die Argonics GmbH, das Ingenieurbüro Kauppert GmbH, die in – innovative navigation GmbH sowie die Bundesanstalt für Wasserbau. Das Aachener Teilprojekt befasst sich mit der satellitenbasierten Lokalisierung der Binnenschiffe und darauf basierend mit den Algorithmen für eine interaktive Routenplanung. Mit diesen Funktionen wird der Automatisierungsgrad erhöht und der Schiffführer entlastet. Im Idealfall kann ein Operator mehrere Schiffe gleichzeitig von einer Leitstelle oder von zu Hause steuern. Mit einem Demonstratorschiff werden die Entwicklungen auf dem Dortmund-Ems-Kanal zwischen dem Hafen Dortmund und der Schleuse Waltrop getestet.

Julius Kühn-Preis für Caspar Langenbach

Dr. Caspar Langenbach vom Lehr- und Forschungsgebiet Biochemie und Molekularbiologie der Pflanzen hat für seine Forschung zur genetischen und biologischen Kontrolle des Asiatischen Sojabohnenrostpilzes den Julius Kühn-Preis 2020 der Deutschen Phytomedizinischen Gesellschaft erhalten. Dieser wird an Wissenschaftlerinnen und Wissenschaftler unter 40 Jahren verliehen, die sich in der Grundlagenforschung auf dem Gebiet der Phytomedizin verdient gemacht haben. Langenbach studierte Biologie an der RWTH, wo er 2013 auch promovierte. Zurzeit leitet er die Arbeitsgruppe „Agrarbiotechnologie“ des Lehr- und Forschungsgebietes Biochemie und Molekularbiologie der Pflanzen unter Leitung von Professor Uwe Conrath.

Hans Hermann Voss-Stiftung fördert Abschlussarbeiten

Die Hans Hermann Voss-Stiftung unterstützt Masterarbeiten von Studierenden, die in einem der Profildbereiche der Hochschule forschen, und mit ihrer Arbeit zur Lösung anwendungsbezogener Fragestellungen beitragen. Mit 159.000 Euro finanziert die Stiftung

insbesondere die Materialkosten für Prototypen. Die Ergebnisse der ersten Förderperiode werden im Mai 2021 auf einem gemeinsamen Symposium mit der Hans Hermann Voss-Stiftung präsentiert.

Fabian Kießling geehrt

Professor Fabian Kießling, Inhaber des Lehrstuhls für Experimentelle Molekulare Bildgebung, wurde zum Fellow der World Molecular Imaging Society ernannt. Die Gesellschaft würdigt damit Kießlings wegweisende Beiträge im Bereich der molekularen Bildgebung. Fabian Kießling studierte Medizin an der Ruprecht-Karls-Universität Heidelberg und promovierte 2001 im Fach Innere Medizin. Anschließend arbeitete er im Deutschen Krebsforschungszentrum, bevor er 2003 in die Abteilung für Onkologie der Thoraxklinik Heidelberg wechselte. 2008 folgte er dem Ruf der RWTH, zugleich wurde er Direktor des Lehr- und Forschungsgebiets Molekulare Bildgebung am Helmholtz-Institut für Biomedizinische Technik.

Förderung für Batterieforschung

Im Rahmen des Dachkonzepts „Forschungsfabrik Batterie“ fördert das Bundesministerium für Bildung und Forschung vier Batterie-Kompetenzcluster, die RWTH ist an drei Clustern beteiligt. Das Ministerium will die Batterieforschung in Deutschland und den Transfer in die Industrie konsequent ausbauen. Die neuen Kompetenzcluster bauen auf den vorhandenen Strukturen des Dachkonzeptes auf.

Der mit rund 30 Millionen Euro geförderte Cluster „Intelligente Batteriezellproduktion“ (InZePro) erforscht die Erhöhung und Flexibilisierung der Produktivität der Zellproduktion. Das soll durch eine ganzheitliche Optimierung des Produktionssystems unter Einsatz von Lösungen der Industrie 4.0 erreicht werden. Schwerpunkte sind eine innovative agile Anlagentechnik, Digitalisierung, Künstliche Intelligenz in der Produktion und virtuelle Produktionssysteme. Professor Achim Kampker vom Lehrstuhl für Production Engineering of E-Mobility Components koordiniert das Projekt. Beteiligt sind die TU München, das Karlsruher Institut für Technologie und die TU Braunschweig.

Der Kompetenzcluster „Recycling/Grüne Batterie“, kurz greenBatt, wird ebenfalls mit rund 30 Millionen Euro unterstützt. Professor Bernd Friedrich vom Institut für Metallurgische Prozesstechnik und Metallrecycling untersucht gemeinsam mit der TU Braunschweig

und dem Fraunhofer-Institut für Keramische Technologien und Systeme die systematische Gestaltung des Batterielebenszyklus. Weitere Schwerpunkte sind die Berücksichtigung und Weiterentwicklung effizienter Recyclingtechnologien und die Integration von rückgewonnenen Materialien in die Batteriezellproduktion mit dem Ziel, Stoffkreisläufe zu schließen. Mit rund 20 Millionen Euro wird der Cluster „Batterienutzungskonzepte“ (BattNutzung) gefördert. Ziel ist, ein tiefgehendes Verständnis von Batteriezuständen und -verhalten zu entwickeln, um zu entscheiden, wann die Zweitnutzung von Batteriespeichern möglich und für welche Anwendungen sinnvoll ist. Professor Dirk Uwe Sauer, Institut für Stromrichtertechnik und Elektrische Antriebe, übernimmt die Sprecherfunktion, beteiligt sind die TU München und das Fraunhofer-Institut für Siliziumtechnologie.

RWTH in Rankings

Jährlich rankt die National Taiwan University die 800 Top-Universitäten der Welt in sechs Wissenschaftsbereichen sowie in 24 einzelnen Fächern. Für das Ranking werden bibliometrische Indikatoren wie internationale wissenschaftliche Publikationen ausgewertet und nach wissenschaftlicher Produktivität, Exzellenz und Einfluss in der Forschung gewichtet. Auf nationaler Ebene schneidet die RWTH im Ranking sehr erfolgreich ab: In 18 von 24 Fächern ist sie unter den stärksten Hochschulen des Landes gelistet. In den Fächern Bauingenieurwesen, Chemieingenieurwesen und Maschinenbau belegt die RWTH den ersten Platz. Mit einer Zweitplatzierung im Fach Elektrotechnik sowie Drittplatzierungen in Chemie, Informatik und Mathematik werden weitere Top-Ergebnisse im bundesweiten Vergleich erreicht.

U-Multirank – initiiert und finanziert von der Europäischen Kommission – bescheinigt der RWTH in vielen Kriterien wieder die Bestnote. Besonders gut schneidet die RWTH in den Bereichen „Research“, „Knowledge Transfer“ und „International Orientation“ ab. In der Gesamtbetrachtung erhielt die RWTH bei 23 von 32 bewerteten Indikatoren die Noten „sehr gut“ und „gut“. Über alle Kriterien hinweg erhält die Hochschule eine Durchschnittsnote von 2,06 und kann sich damit im Vergleich zum Vorjahr leicht verbessern. Besonders gut fällt die Bewertung der Forschungsleistung und beim Wissenstransfer aus. Bei den Kriterien wie privaten Forschungsgeldern, Spin-offs, in Patenten zitierten Publikationen und gemeinsamen Veröffentlichungen mit

Partnern aus der Industrie wird die Note 1 vergeben.

Top-Gründeruni

Laut Bundesministerium für Bildung und Forschung gehört die RWTH zu den Top-Hochschulen für Förderungen von Gründungsvorhaben im Förderjahr 2019/20. So belegen RWTH und Uniklinik RWTH Aachen im Rahmen des EXIST-Programms mit insgesamt 19 bewilligten Förderungen den zweiten Platz bundesweit. Mit drei geförderten Forschungstransfer-Projekten und 16 bewilligten Gründerstipendien ist die RWTH die einzige NRW-Hochschule unter den Top 10.

RWTH koordiniert Projekt „AquaSPICE“

Professor Thomas Wintgens vom Institut für Siedlungswasserwirtschaft koordiniert das Projekt „AquaSPICE – Advancing Sustainability of Process Industries through Digital and Circular Water Use Innovations“. Die Europäische Union finanziert die Forschungsarbeiten zur industriellen Wasserwirtschaft in den nächsten 3,5 Jahren im Rahmen des Forschungs- und Innovationsprogramms H2020 mit elf Millionen Euro. Beteiligt sind 29 Partner aus zwölf Ländern.

Die Industrie ist einer der großen Wassernutzer in Europa, die Wiederverwendung von bereits genutztem Wasser ist von wachsender Bedeutung. Im Projekt soll durch die Verbindung von Aufbereitungsverfahren, Sensortechnologien und digitalen Prozessen die Schließung der Wasserkreisläufe vorangetrieben werden. Daher engagieren sich Vertreter mittlerer und großer Industriebetriebe aus den Bereichen Nahrungsmittelherstellung, Chemie, Mineralölverarbeitung sowie Abfallverwertung ebenso wie Forschungsinstitutionen und Softwareanbieter. In sechs Fallstudien werden Wasserströme für die Produktionsprozesse mit Ansätzen der Kreislaufwirtschaft betrachtet und mit verschiedenen Behandlungsverfahren im Pilotmaßstab aufbereitet. Eingesetzt werden biologische und physikalisch-chemische Verfahren, unter anderem auch Membranprozesse. Der Zustand von Wasserressourcen und Aufbereitungsprozessen soll mit Sensorsystemen in Echtzeit erfasst und mit Hilfe einer digitalen Plattform optimiert werden.

Highly Cited Researchers

Drei Wissenschaftler sind in der Liste der „Highly Cited Researchers“ vertreten. Zu den weltweit am häufigsten zitierten Forschenden gehören Fabian Kießling und Twan Lammers

sowie Ralph Panstruga. Kießling ist Inhaber des Lehrstuhls für Experimentelle Molekulare Bildgebung, Lammers betreut das Lehr- und Forschungsgebiet für Nanomedizin und Theranostik. Beide waren bereits 2019 als Highly Cited Researcher gelistet. Panstruga, Lehr- und Forschungsgebiet Molekulare Zellbiologie der Pflanzen, wurde bereits in den Jahren 2014, 2015 und 2019 in die Liste aufgenommen.

RWTH im Verbund Nationales Hochleistungsrechnen

Die RWTH gehört künftig dem von Bund und Ländern geförderten Verbund für das Nationale Hochleistungsrechnen an. Die Gemeinsame Wissenschaftskonferenz empfahl, die RWTH und die TU Darmstadt nach deren gemeinsamer Antragstellung in das Konsortium aufzunehmen. Bundesweit werden erstmalig acht Rechenzentren von Hochschulen aufgenommen. Grundlage ist die Entscheidung von Bund und Ländern, neben Forschungsbauten und Großgeräten das Hochleistungsrechnen an Hochschulen mit 62,5 Millionen Euro jährlich über einen Zeitraum von zehn Jahren zu fördern.

Mehrere Rechenzentren von überregionaler Bedeutung werden künftig in einer zukunfts-fähigen Struktur zusammengefasst. Bund und Länder wollen nicht nur Investitionskosten, sondern auch Betriebs- und Personalkosten übernehmen. So können neben der HPC-Infrastruktur auch Unterstützungsdienstleistungen für die Nutzenden angeboten werden. Darauf aufbauend bündeln die Standorte Aachen und Darmstadt in dem „Nationalen Hochleistungsrechenzentrum for Computational Engineering Sciences“ (NHR4CES) ihre bestehenden Stärken in High Performance Computing-Anwendungen, Algorithmen und Methoden sowie bei der effizienten Bereitstellung und Nutzung von HPC-Hardware. Zudem fördert das Mentoring-Programm von NHR4CES gezielt den wissenschaftlichen Nachwuchs.

Nationale Forschungsdateninfrastruktur

Die Gemeinsame Wissenschaftskonferenz hat neun Konsortien für den Aufbau einer Nationalen Forschungsdateninfrastruktur (NFDI) bewilligt. An den Konsortien „NFDI4Ing“, „NFDI4Cat“ und „NFDI4Chem“ ist die RWTH beteiligt, bei „NFDI4Ing“ sogar als Konsortialführerin. Im Rahmen einer Nationalen Forschungsdateninfrastruktur sollen für das deutsche Wissenschaftssystem Standards im Datenmanagement gesetzt und ein digitaler, regional verteilter und vernetzter Wissensspeicher Forschungsdaten nachhaltig sichern und nutzbar machen. Die NFDI wird über einen Zeitraum von drei Jahren in drei Förderrunden aufgebaut. In jeder Runde können neue Konsortien aufgenommen werden. Für Aufbau und Förderung werden Bund und Länder jährlich bis zu 90 Millionen Euro bereitstellen.

NFDI4Ing

NFDI4Ing hat sich zum Ziel gesetzt, ein Forschungsdatenmanagement zu etablieren, das zugleich die Transparenz und Nachvollziehbarkeit ingenieurwissenschaftlicher Forschungsergebnisse stärkt. Es ist der einzige geförderte Verbund im Bereich der Ingenieurwissenschaften und wurde von der RWTH und der Technischen Universität Darmstadt initiiert und durch die Aufnahme weiterer Partner ausgebaut. Drei Aspekte werden fokussiert: die Verbesserung der Ausbildung in der Datenkompetenz, die Entwicklung eines Governance-Konzepts für den Umgang mit Forschungsdaten und die Sicherstellung einer breiten Verfügbarkeit von Technologien und Services für maschinenlesbare Daten. Sprecher ist Professor Robert Schmitt vom Lehrstuhl für Fertigungsmesstechnik und Qualitätsmanagement.

NFDI4Chem

NFDI4Chem soll alle wissenschaftlichen Disziplinen der Chemie repräsentieren. Die Idee ist die Digitalisierung aller wichtigen Schritte in der chemischen Forschung, um Forschende bei der Sammlung, Speicherung, Verarbeitung, Analyse, Offenlegung und Wiederverwendung ihrer Daten zu unterstützen. Ziel ist dabei die Erstellung eines ganzheitlichen Konzeptes. Als Maßnahmen fungieren die Förderung von Open Science und Forschungsdatenmanagement in Übereinstimmung mit den FAIR-Prinzipien (Findable, Accessible, Interoperable und Re-usable). Das Konsortium wird von der Gesellschaft Deutscher Chemiker (GDCh), der Deutschen

Bunsen-Gesellschaft für physikalische Chemie (DBG) und der Deutschen Pharmazeutischen Gesellschaft (DPHG) unterstützt. Professorin Sonja Herres-Pawlis vom Institut für Anorganische Chemie ist Co-Sprecherin.

NFDI4Cat

NFDI4Cat besteht aus Experten der Bereiche Katalysatorforschung, Reaktions- und Verfahrenstechnik sowie der Informationstechnologie. Die Katalyse ist eine der Kerntechnologien, um die globalen Herausforderungen des Klimawandels sowie der nachhaltigen Energie- und Materialversorgung zu lösen. Dazu ist ein grundlegender Wandel in der Katalysatorforschung, der chemischen Verfahrenstechnik und der Prozesstechnologie erforderlich. Eine Herausforderung ist dabei die Zusammenführung der verschiedenen Disziplinen. In diesem Zusammenhang bedarf es neben einem gemeinsamen Verständnis der Katalysatorwissenschaften und der entzirkelnden Entwicklung universeller Leitprinzipien auch der Vereinheitlichung von Arbeitsabläufen. Ergebnisse müssen zudem allen Forschenden gleichermaßen zugänglich sein, dazu ist die Schaffung hochleistungsfähiger Informationsarchitekturen erforderlich. Diese ermöglichen die Speicherung, den Austausch und die Analyse von Daten unter Verwendung der Werkzeuge der künstlichen Intelligenz. Softwarepakete und -tools erlauben die Erweiterung der Theorie um Bereiche wie Simulation und experimentelle Studien in Design, Charakterisierung und Kinetik von Katalysatoren und den verbundenen technischen Aspekten. Die RWTH fungiert als co-antragstellende Einrichtung.



Im Jahr 2020 wurde erstmals eine auf Quantenphysik basierende Technologie, entwickelt in der Schweiz, in ein Smartphone eingebaut. Das soll die Datensicherheit im Mobilfunk erheblich verbessern.

Foto: Peter Winandy

Impressum

Herausgegeben im Auftrag des Rektors
der RWTH Aachen
Dezernat 3.0 – Presse und Kommunikation
Templergraben 55
52056 Aachen
Telefon +49 241 80 - 93687
pressestelle@rwth-aachen.de
www.rwth-aachen.de

Verantwortlich:
Renate Kinny

Redaktion:
Angelika Hamacher

Titelbild:
Peter Winandy, Aachen

Anzeigen:
Medienhaus Aachen GmbH
Dresdener Str. 3, 52068 Aachen
Geschäftsführer: Andreas Müller
Leiter Werbemarkt: Jürgen Carduck
(verantwortlich für Anzeigen i.S.d. § 8 Abs. 2
Landespressegesetz NRW)

Martina Welz Assistenz Anzeigenleitung
Tel. 0241/5101 611
E-Mail: martina.welz@medienhausaaachen.de

Gestaltung:
Kerstin Lünenschloß, Aachen

Druck:
image Druck + MEDIEN GmbH, Aachen

Das Wissenschaftsmagazin RWTH THEMEN
erscheint einmal pro Semester.

Nachdruck einzelner Artikel, auch auszugs-
weise, nur mit Genehmigung der Redaktion.

Für den Inhalt der Beiträge sind die Autoren
verantwortlich.

ISSN-Nummer 0179-079X

„Als Ingenieur arbeite ich bei Viessmann gemeinsam mit IT Experten an nahtlosen Klima- und Energielösungen für zukünftige Generationen.

So entstehen für unsere Kunden innovative Produkte und für mich Chancen zur kontinuierlichen Weiterentwicklung.“

Dr.-Ing. Moritz Lauster,
Senior Product Manager und
RWTH Aachen Alumnus



Bewirb dich jetzt
viessmann.family/jobs