DAGSTUHL
REPORTS

Volume 7, Issue 2, February 2017

ISSN 2192-5283

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at
http://www.dagstuhl.de/dagpub/2192-5283

Publication date
September, 2017

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at `http://dnb.d-nb.de`.

Digital Object Identifier: 10.4230/DagRep.7.2.i

Aims and Scope
The periodical *Dagstuhl Reports* documents the program and the results of Dagstuhl Seminars and Dagstuhl Perspectives Workshops.
In principal, for each Dagstuhl Seminar or Dagstuhl Perspectives Workshop a report is published that contains the following:
- an executive summary of the seminar program and the fundamental results,
- an overview of the talks given during the seminar (summarized as talk abstracts), and
- summaries from working groups (if applicable).

This basic framework can be extended by suitable contributions that are related to the program of the seminar, e. g. summaries from panel discussions or open problem sessions.

Contact
Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de
http://www.dagstuhl.de/dagrep

Report from Dagstuhl Seminar 17061

Wildly Heterogeneous Post-CMOS Technologies Meet Software

Edited by

Jerónimo Castrillón-Mazo[1], Tei-Wei Kuo[2], Heike E. Riel[3], and Matthias Lieber[4]

1 **TU Dresden, DE,** jeronimo.castrillon@tu-dresden.de
2 **National Taiwan University – Taipei, TW,** ktw@csie.ntu.edu.tw
3 **IBM Research Zurich, CH,** hei@zurich.ibm.com
4 **TU Dresden, DE,** matthias.lieber@tu-dresden.de

Abstract

The end of exponential scaling in conventional CMOS technologies has been forecasted for many years by now. While advances in fabrication made it possible to reach limits beyond those predicted, the so anticipated end seems to be imminent today. The main goal of the seminar 17061 "Wildly Heterogeneous Post-CMOS Technologies Meet Software" was to discuss bridges between material research, hardware components and, ultimately, software for information processing systems. By bringing together experts from the individual fields and also researchers working interdisciplinarily across fields, the seminar helped to foster a mutual understanding about the challenges of advancing computing beyond current CMOS technology and to create long-term visions about a future hardware/software stack.

Seminar February 5–10, 2017 – http://www.dagstuhl.de/17061

1998 ACM Subject Classification B.7.1 [Integrated Circuits]: Types and Design Styles – Advanced technologies, Algorithms implemented in hardware, Memory technologies, C.1.3 [Processor Architectures]: Other Architecture Styles – Analog computers, Data-flow architectures, Heterogeneous (hybrid) systems, Neural nets, C.3 [Special-purpose And Application-based Systems]: Real-time and embedded systems, Signal processing systems, C.4 [Performance Of Systems]: Design studies, Modeling techniques, D.3.2 [Programming Languages]: Language Classifications – Data-flow languages, Specialized application languages, D.3.4 [Programming Languages]: Processors – Code generation, Compilers, Run-time environments, D.4.7 [Operating Systems]: Organization and Design

Keywords and phrases 3D integration, compilers, emerging post-CMOS circuit materials and technologies, hardware/software co-design, heterogeneous hardware, nanoelectronics, neuromorphic computing, operating systems, programming languages, runtime support, system architecture

Digital Object Identifier 10.4230/DagRep.7.2.1

1 Executive Summary

Jerónimo Castrillón-Mazo
Tei-Wei Kuo
Heike E. Riel
Sayeef Salahuddin
Matthias Lieber

Topic and Structure

The end of exponential scaling in conventional CMOS technologies has been forecasted for many years by now. While advances in fabrication made it possible to reach limits beyond those predicted, the so anticipated end seems to be imminent today. An indication of this is the research boom, both in academia and industry, in emerging technologies that could complement or even replace CMOS devices. Examples for such emerging technologies include tunnel FETs, nonvolatile memories such as magnetoresistive RAM, 3D integration, carbon nanotube transistors, and graphene.

The main goal of this seminar was to discuss bridges between material research, hardware components and, ultimately, software for information processing systems. Given a new class of wildly heterogeneous systems that integrate different technologies, we want to reason about enabling hardware and software abstractions, from languages and system-software down to hardware mechanisms. The challenge of realizing an efficient wildly heterogeneous system can only be tackled by employing holistic and synergistic approaches in an interdisciplinary environment. By bringing together experts from the individual fields and also researchers working interdisciplinarily across fields, the seminar helped to foster a mutual understanding about the challenges of advancing computing beyond current CMOS technology and to create long-term visions about a future hardware/software stack.

The seminar was structured around four partially overlapping areas, namely: (i) far-fetched materials and physics such as spin, nanomagnets, phase transition, and correlated phenomena, (ii) near future materials (and software) such as phase-change memory, nanowires, nanotubes, and neuromorphic devices, (iii) low-level software layers for new technologies such as operating systems, runtime support, middleware, and HW/SW-co-designed firmware, and (iv) upper software layers such as new programming/specification languages, models, and software synthesis.

Important questions addressed by the seminar included:

- Materials/Devices: What are the current status and the roadmap of post-CMOS materials and technologies? What will be the expected characteristics of the new devices? Will new technologies enable a fundamentally different computing paradigm, e. g., beyond von Neumann? What are the challenges for proper benchmarking of different technologies?
- Hardware/Software Stack: How much of the hardware's heterogeneity and its characteristics should be exposed to programmers? How general may be a programming model/language for future (yet unknown) hardware? How to make software adapt itself to hardware with fluctuating resources? Which new applications can be enabled by emerging materials and technologies and what needs to be done at the software layers to make them viable?
- Analysis: How can we model the interactions across the layers of the hardware/software stack? What kind of formal operational models and analysis methods are needed for evaluating heterogeneous systems? Can system-level analysis of new technologies give insights to material scientists, disrupting the otherwise incremental innovation paradigm?

Main Conclusions

Summary

There will probably be no CMOS replacement for chips with billions of transistors in the next 20 years, but architectural advances at various levels (such as 3D transistors, 3D integration of memory and logic, specialization, and reconfigurability) will lead to performance improvements despite the scaling limitations of planar CMOS technology. New non-volatile memories (e. g., spin-based) bear the potential to radically change various areas of computing, such as data-intensive processing and neuromorphic computing. New hardware architectures will need rethinking today's software stack and our widely used programming models. Finally, even though some post-CMOS technologies will not replace high-end CMOS transistors, there is great potential in new, yet unknown, applications. Applications, backed by a strong commercial demand, will give some technologies the push to become viable. Examples are radio-frequency for carbon nanotubes, graphene based sensors, organic low-cost transistors for wearables, and memristors for neuromorphic computing.

Post-CMOS logic for compute-intensive applications

Currently, there is no alternative to CMOS on the horizon to realize logic for large von Neumann computing, due to lower projected performance and/or yield challenges. Candidates discussed on the seminar have been: tunnel FETs, III-V, 2D materials such as graphene, CNTs, or spintronics. This means that general purpose and high-performance computing will most probably be based on CMOS in the medium term. To workaround the CMOS scaling problem, architectural specialization will gain more and more importance leading to general purpose computing systems with (various) specialized accelerators. We already find them today in, e. g., mobile devices or GPU high-performance computing accelerators. Additionally, reconfigurable logic, such as FPGAs, and application-specific circuits have a high potential for performance gains. However, it is a big challenge to program such heterogeneous systems. Work towards solutions based on dataflow programming, memory access patterns, skeletons, and domain-specific languages have been discussed at the seminar. Additionally, operating system might need to adapt to allow, for example, accelerators to perform system calls.

Emerging memory technologies

In near future, new non-volatile memories will be available that could unify RAM and permanent storage, including MRAM and RRAM. While these could provide huge benefits for memory-intensive applications, the implications on architecture and software stack are not yet clear. For example, what will be the role of the file system in such an architecture? And how to deal with security aspects when every bit in RAM is permanent? Looking further into the future, the spin-based, non-volatile racetrack memory has the potential to compete with SRAM in terms of performance, while consuming considerably less energy. High-performance and energy-efficient non-volatile memories will also be important for neuromorphic devices.

Going 3D

3D integration enables the integration of heterogeneous technologies for logic, memory, communication, and sensing on a single chip. At the transistor level, 3D corrugated transistors were discussed as a promising direction to keep reducing the footprint while avoiding short-channel effects. Advancing today's die stacking technology through fine-grained vias linking the layers, will provide a substantial improvement for latencies and bandwidths in the systems. Bringing memory closer to logic will lower the memory wall (or even lead to a breakdown?). This means that many existing applications could be compute bound (again) and the processor architecture could be potentially simplified by removing the overhead that was added to workaround the memory wall, such as big caches and prefetchers, making place for additional compute units. In this optimistic scenario, general-purpose computing would receive a great (one-time) performance boost. For compilers and applications, we would have to rethink our way of optimizing code.

Computing beyond von Neumann

Architectural approaches beyond von Neumann were also discussed to speedup specific applications. Examples were neuromorphic computers, analog circuits, and dataflow machines. Of course these approaches cannot replace general purpose processors completely. A possible future architecture would combine classical von Neumann processors with non-von Neumann accelerators (on the same chip) to enable mixed programming. The recent industry adoption of machine learning drives the need for neuromorphic computers. While these systems already outperform general purpose processors today, new technologies such as non-volatile memories and analog spintronics promise even greater gains. Promising analog circuits were shown to perform well for concrete NP-complete problems such as SAT and graph coloring. Along these lines, a theoretical framework was introduced that may serve to abstractly compare the asymptotic energy efficiency between the analog and the digital realizations of a system. Finally, dataflow machines where discussed that stream data directly between computational units without the overhead of registers and caches, thereby removing the “Turing tax”.

Special applications

Some of the materials considered in the seminar are very likely not able to compete with CMOS for logic, but have strengths in other electronic application areas such as sensors, radio frequency, and displays. Carbon nanotubes and graphene are promising materials for high-frequency antennas required for upcoming wireless communication systems. In the particular case of Graphene, it seems that the initial technological hype has passed, and engineering has taken over to produce new clever devices (e. g., nano-membranes for sensing). Organic electronics are already commercially available in displays and OLEDs. Their distinct features of flexibility, low production cost (printed electronics), and biodegradability could potentially open completely new application areas for logic, but not at comparable speed and efficiency to CMOS. These devices have also been deemed important for bio-compatibility. However, there is a long road ahead for testing and certifying actual devices in living tissue, which is not a trivial task, considering the wealth of molecules being investigated in this domain.

Co-design and design space exploration

Proper hardware/software co-design will be very important to achieve performance gains given the limits of CMOS and the prospective wildly heterogeneous and/or application-specific computing systems. Given a specific application problem, numerous implementation alternatives, from the algorithm down to the hardware architecture and technologies, might be feasible. Tools that help developers navigating the huge design space (e. g., using modeling and benchmarking techniques) and automate an efficient implementation as much as possible are needed. It appears to be that the large part of the software is less flexible than the hardware and much work has to be done to make software future-proof.

2 Table of Contents

3 Overview of Talks

3.1 Stochastic power management in energy harvesting systems

Rehan Ahmed (ETH Zürich, CH)

Joint work of Rehan Ahmed, Bernhard Buchlil, Pratyush Kumar, Lothar Thiele

Ambient energy harvesting has been shown to have significant potential in increasing the lifetime of sensor motes and IoT devices. However, energy harvesting sources are variable in nature, and good prediction/power management strategies need to be designed so that the systems powered by them do not encounter battery depletion. In this work, we present a formal study on optimizing the energy consumption of energy harvesting embedded systems. To deal with the uncertainty inherent in these systems, we have developed a Stochastic Power Management (SPM) scheme, that builds statistical models of harvestable energy based on historical data, and uses these models to design an energy consumption profile. The proposed scheme, maximizes the minimum energy consumption over all time intervals, while giving probabilistic guarantees on not encountering battery depletion. We also present results of experimental evaluation. Through the results, we quantitatively establish that the proposed solution is highly effective at providing a guaranteed minimum service level.

3.2 Neuromorphic computing with Non-Volatile Memories

Stefano Ambrogio (IBM Almaden Center – San Jose, US)

Neuromorphic computing stands as an innovative solution for targeting tasks which are easily solved by the human brain, but that require high computational resources on current Von-Neumann computers. This talk presents a brief overview of the main research branches employing Non-Volatile Memories (NVM) as the synaptic element in neural networks for Machine Learning [1]. This research field has gained an increasing interest in the last years due to the performance opportunities that NVM could potentially provide, outperforming nowadays GPUs and CPUs [1, 2].

First, the talk targets fully connected neural networks with Phase Change Memory, trained with the backpropagation algorithm. After introducing the working principle, recent results and comparison between devices used in analog or binary modes are provided [1, 3]. Then, the talk shows some networks trained with the Spike-Timing-Dependent-Plasticity biological protocol [4, 5], underlining the differences with the backpropagation algorithm and the need for extensive global studies in this field. Finally, the impact of device non-idealities on both backpropagation and STDP networks and algorithms is analyzed and some solutions are provided.

References

1 G. W. Burr et al. *Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165 000 Synapses) Using Phase-Change Memory as the Synaptic Weight Element.* IEEE Trans. Elec. Dev., 62(11):3498–3507, 2015.

2 G. W. Burr et al. *Large-scale neural networks implemented with non-volatile memory as the synaptic weight element: Comparative performance analysis (accuracy, speed, and power).* In IEDM 2015.
3 S. Yu et al. *Binary neural network with 16 Mb RRAM macro chip for classification and online training.* In IEDM 2016.
4 S. Kim et al. *NVM neuromorphic core with 64k-cell (256-by-256) phase change memory synaptic array with on-chip neuron circuits for continuous in-situ learning.* In IEDM 2015.
5 S. Ambrogio et al. *Neuromorphic learning and recognition with one-transistor-one-resistor synapses and bistable metal oxide RRAM.* IEEE Trans. Elec. Dev., 63(4):1508–1515, 2016.

3.3 M3: Integrating Arbitrary Compute Units as First-class Citizens

Nils Asmussen (TU Dresden, DE)

Joint work of Nils Asmussen, Marcus Völp, Benedikt Nöthen, Hermann Härtig, Gerhard Fettweis
Main reference N. Asmussen, M. Völp, B. Nöthen, H. Härtig, G. Fettweis, "M3: A Hardware/Operating-System Co-Design to Tame Heterogeneous Manycores", in Proc. of the 21st Int'l Conf. on Architectural Support for Prog. Lang. and Oper. Systems (ASPLOS 2016), pp. 189–203, ACM, 2016.
URL http://dx.doi.org/10.1145/2872362.2872371

We are currently observing a trend towards more heterogeneous systems in order to meet the desired performance and energy efficiency. For example, DSPs, FPGAs and special purpose accelerators are employed next to general purpose cores. However, current operating systems are relying on processor features such as user/kernel mode and memory management units for protection and access to operating system services. These features are not necessarily available on all compute units (CUs), preventing an integration of arbitrary CUs as first-class citizens.

I will present a hardware/software co-design, consisting of a new hardware component and an operating system based on it. By introducing a common interface for all CUs, arbitrary CUs can be integrated as first-class citizens, where untrusted code can run on all CUs and all CUs can access operating system services such as file systems or network stacks.

3.4 Exploring Performance Portability using Memory-Oriented Programming Models

Tal Ben-Nun (The Hebrew University of Jerusalem, IL)

As heterogeneous computing architectures become ubiquitous, application programming complexity is increasing beyond the skill-set of the average scientist. Thus, it is imperative to devise a unified programming environment that enables efficient utilization of the underlying computational resources, without sacrificing simplicity. The talk will present a programming model that tackles one of the fundamental aspects of computing – memory access – which often recurs as a performance bottleneck in parallel applications. The talk will show that by categorizing algorithm inputs and outputs into access patterns, a wide variety of programs can be automatically optimized for various architectures and partitioned across multiple devices. Using the memory-oriented representation, both processing architectures (e. g., CPU, GPU, FPGA) and memory architectures (e. g., stacked memory, ReRAM) can potentially be utilized to their full extent. The presented memory-oriented programming model currently exhibits state-of-the-art performance on nodes with multiple GPUs and irregular algorithms, facilitating the development of efficient applications on architectures that range from mobile devices to supercomputers.

3.5 A hardware/software stack for emerging systems

Jerónimo Castrillón-Mazo (TU Dresden, DE)

This talk introduces the German cluster of excellence "Center for Advancing Electronics Dresden" (cfaed, http://cfaed.tu-dresden.de), which looks into a set of promising technologies that may augment or replace CMOS. Given a new class of wildly heterogeneous systems that integrate different technologies, the Orchestration sub-project of cfaed aims at devising hardware and software abstractions that would allow programming such complex systems [6]. Abstractions include those typically found in computing systems, ranging from hardware mechanisms up to software engineering approaches. These abstractions are paired with formal modelling for quantitative analysis, aiming at tradeoff analysis in heterogeneous systems.

We discuss general hardware mechanisms for isolation on tile-based systems, exemplarily demonstrated in the Tomahawk multicore platform [1]. We argue that tile-based systems offer a well-suited architectural template for integrating components implemented in different technologies. At the hardware level, components must only agree on the interfacing to the on-chip network via routers that provide isolation at the hardware level. We describe the M3 capability-based operating system (OS) [2] which builds on this hardware interface. With a micro-kernel approach, M3 provides access to system resources for tiles that cannot run a full-fledged OS (see abstract 3.3 by Nils Asmussen). As programming abstraction, we use dataflow programming models, architecture models and compilers to automatically generate low-level code for heterogeneous multi-cores [4]. Finally, for formal trade-off analysis, we have developed new theory to handle multiple objective functions and resolve nondeterministic choices in an optimal way [3].

In this talk, we report on early results of deploying the abstractions on a heterogeneous CMOS platform and an effort to bring up a system simulator that allows integrating models of components on Post-CMOS technologies. We briefly discuss promising architectural options with reconfigurable 1D transistors (e. g., with silicon nano-wires [5]). Finally, we share our experience when trying to bridge the broad interdisciplinary gap between material and computer scientists.

References

1 O. Arnold et al. *Tomahawk: Parallelism and heterogeneity in communications signal processing MPSoCs*. ACM Transactions on Embedded Computing Systems, 13(3s):107, 2014.
2 N. Asmussen et al. *M3: A hardware/operating-system co-design to tame heterogeneous manycores.* In ASPLOS 2016.
3 C. Baier et al. *Probabilistic model checking and non-standard multi-objective reasoning.* In FASE 2014.
4 J. Castrillon and R. Leupers. *Programming Heterogeneous MPSoCs: Tool Flows to Close the Software Productivity Gap.* Springer, 2014.

5 M. Raitza et al. *Exploiting transistor-level reconfiguration to optimize combinational circuits.* In DATE 2017.
6 M. Völp et al. *The Orchestration Stack: The Impossible Task of Designing Software for Unknown Future Post-CMOS Hardware.* In PMES at SC 2016.

3.6 Random Thoughts/Examples about Neuromorphic Computing and Emerging Devices

Yiran Chen (Duke University – Durham, US)

Joint work of Yiran Chen, Hai (Helen) Li, Qing Wu, Wei Wen, Beiye Liu, Chaofei Yang, Miao Hu

Human brain is the most sophisticated organ that nature ever builds. Building a machine that can function like a human brain, indubitably, is the ultimate dream of a computer architect. Although we have not yet fully understood the working mechanism of human brains, the part that we have learned in past seventy years already guided us to many remarkable successes in computing applications, e. g., artificial neural network and machine learning. The recently emerged research on "neuromorphic computing", which stands for hardware acceleration of brain-inspired computing, has become one of the most active areas in computer engineering. Our presentation starts with a background introduction of neuromorphic computing, followed by some design examples of hardware acceleration schemes of learning and neural network algorithms on IBM TrueNorth and memristor-based computing engines. At the end, we will share our prospects on the future technology challenges and advances of neuromorphic computing.

References

1 W. Wen et al. *A new learning method for inference accuracy, core occupation, and performance co-optimization on TrueNorth chip.* In DAC 2016.
2 M. Hu et al. *Memristor Crossbar-Based Neuromorphic Computing System: A Case Study.* IEEE Trans. Neural Netw. Learning Syst., 25(10):1864–1878, 2014.
3 C. Yang et al. *Security of neuromorphic computing: thwarting learning attacks using memristor's obsolescence effect.* In ICCAD 2016.

3.7 Bridging the gap between single device fabrication and system design for emerging device technologies

Martin Claus (TU Dresden, DE)

Main reference M. Schröter, M. Claus, P. Sakalas, M. Haferlach, D. Wang, "Carbon Nanotube FET Technology for Radio-Frequency Electronics: State-of-the-Art Overview", IEEE J-EDS, 1(1):9–20, 2013.
URL http://dx.doi.org/10.1109/JEDS.2013.2244641

Research on emerging electronics is in many cases restricted to single-device fabrication as a proof of concept study for specific two-dimensional (e. g. MoS2) and one-dimensional materials (e. g. CNTs). Depending on the device architecture, these devices capture the functionality (switching and amplification) [1] of incumbent Silicon-based transistors or add functionality such as reconfigurability at the transistor level [2].

However, for evaluating the performance of these materials and new device functionality, circuit and system design studies comprising hundreds or even millions of devices are essential. Since the fabrication of these systems is far beyond the technological possibilities for most emerging technologies, the circuit and system evaluation relies on simulations. Due to the inherent complexity of emerging devices, holistic multi-scale simulations are required [3, 4, 5].

The talk will focus on one-dimensional materials and devices for high-performance computing as well as reconfigurable systems. The significance of holistic multi-scale simulations for technology development as well as circuit and system design based on physics-based compact modeling will be highlighted.

References

1 M. Schröter et al. *Carbon Nanotube FET Technology for Radio-Frequency Electronics: State-of-the-Art Overview*, IEEE Journal of the Electron Devices Society, 1(1):9–20, 2013.
2 G. Darbandy, M. Claus and M. Schröter. *High-Performance Reconfigurable Si Nanowire Field-Effect Transistor Based on Simplified Device Design*, IEEE Transactions on Nanotechnology, 15(2):289–294, 2016.
3 M. Claus et al. *COOS – A wave-function based Schrödinger-Poisson solver for ballistic nanotube transistors*, Journal of Computational Electronics, 13:(2):689–700, 2014.
4 S. Mothes, M. Claus and M. Schröter. *Toward Linearity in Schottky Barrier CNTFETs*, IEEE Transactions on Nanotechnology, 14(2):372–378, 2015.
5 M. Schröter et al. *A Semiphysical Large-Signal Compact Carbon Nanotube FET Model for Analog RF Applications*, IEEE Transactions on Electron Devices, 62(1):52–60, 2015.

3.8 System-Level Design Optimization for Integration with Silicon Photonics

Ayse Coskun (Boston University, US)

System-level design tools and optimization methods are essential for enabling computer engineers experiment with emerging technologies. Similarly, a system-level view is necessary for researchers working with new technologies to work with constraints imposed by different applications or architectures. This talk discusses a cross-layer methodology for designing power-efficient many-core systems with on-chip silicon photonic networks. The proposed methodology enables optimizing the layout [1] or the runtime operation [2] of a target system to reduce the power overhead and/or guardbanding associated with silicon photonics integration on chip.

Through this specific example of integration with silicon photonics, another aim of the talk is to demonstrate a way for enabling early integration of emerging technologies into system design, including when using 2.5D/3D stacking to integrate (broadly) heterogeneous technologies together. The talk also discusses various open design automation and tooling challenges in designing systems with emerging technologies and in heterogeneous system design and runtime management.

References

1 A. K. Coskun et al. *Cross-layer Floorplan Optimization for Silicon Photonic NoCs in Many-core Systems.* In DATE 2016.

2 J. Abellan et al. *Adaptive Tuning of Photonic Devices in a Photonic NoC Through Dynamic Workload Allocation.* IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 36(5):801–814, 2017.

3.9 Matching computer science tools and new technology

Erik P. DeBenedictis (Sandia National Labs – Albuquerque, US)

New devices provide new computational capabilities, yet we will need new computer science tools to exploit the capabilities in algorithms. The talk will highlight two examples, but others can be discussed informally.

The first example is the inability of computational complexity theory to capture the advantage of some new devices. For example, RRAM arrays used in neural networks can have non-unity (energy) cost for scalar multiplication, including unexpectedly low energy. Essentially, the energy of a multiply depends on how the result is used later on. However, the straightforward interpretation of computational complexity theory assumes unit cost for arithmetic (of a given precision). I claim this is why computer science community assumes physicists are in error when they make certain exotic claims about devices. The misunderstanding then blocks development of algorithms using the new devices to best advantage. The proposed resolution is to use a complexity measure for algorithms based on minimum energy in units of kT. I'll present examples of the problem and resolution.

The second issue is the bias towards the von Neumann architecture in computer architecture tools. The HPC computer architecture community uses an iterative process called "codesign" in an attempt to improve architectures for the "post Moore's Law era". This means simulating proposed new architectures against frequently used algorithms or instruction traces, iteratively modifying the architecture to get better performance or energy efficiency. Due to artifacts of the von Neumann architecture in the simulation inputs, if somebody applies codesign to a new non-von Neumann architecture, the feedback process will very quickly restore the architecture to the von Neumann model. I claim this is why we are overwhelmed with minor variants of the von Neumann architecture while not having effective ways to exploit new physics. The remedy is to replace codesign with feedback loop that does not include artifacts of the von Neumann architecture. I will give examples.

3.10 AnyDSL: Building Domain-Specific Languages for Productivity and Performance

Sebastian Hack (Universität des Saarlandes, DE) and Roland Leißa (Universität des Saarlandes, DE)

Joint work of Sebastian Hack, Klaas Boesche, Roland Leißa, Philipp Slusallek
Main reference S. Hack, K. Boesche, R. Leißa, R. Membarth, P. Slusallek, "Shallow embedding of DSLs via online partial evaluation", in Proc. of the 2015 ACM SIGPLAN Int'l Conf. on Generative Prog.: Concepts and Experiences (GPCE 2015), pp. 11–20, ACM, 2015.
URL http://dx.doi.org/10.1145/2814204.2814208

To achieve good performance, programmers have to carefully tune their application for the target architecture. Optimizing compilers fail to produce the "optimal" code because their hardware models are too coarse-grained. Even more, many important compiler optimizations are computationally hard even for simple cost models. It is unlikely that compilers will ever be able to produce high-performance code automatically for today's and future machines.

Therefore, programmers often optimize their code manually. While manual optimization is often successful in achieving good performance, it is cumbersome, error-prone, and unportable. Creating and debugging dozens of variants of the same original code for different target platform is just an engineering nightmare.

An appealing solution to this problem are domain-specific languages (DSLs). A DSL offers language constructs that can express the abstractions used in the particular application domain. This way, programmers can write their code productively, on a high level of abstraction. Very often, DSL programs look similar to textbook algorithms. Domain and machine experts then provide efficient implementations of these abstractions. This way, DSLs enable the programmer to productively write portable and maintainable code that can be compiled to efficient implementations. However, writing a compiler for a DSL is a huge effort that people are often not willing to make. Therefore, DSLs are often embedded into existing languages to save some of the effort of writing a compiler.

In this talk, I will present the AnyDSL framework we have developed over the last three years. AnyDSL provides the core language Impala that can serve as a starting point for almost "any" DSL. New DSL constructs can be embedded into Impala in a shallow way, that is just by implementing the functionality as a (potentially higher-order) function. AnyDSL uses online partial evaluation remove the overhead of the embedding.

To demonstrate the effectiveness of our approach, we generated code from generic, high-level text-book image-processing algorithms that has, on each and every hardware platform tested (Nvidia/AMD/Intel GPUs, SIMD CPUs), beaten the industry standard benchmark (OpenCV) by 10-35 %, a standard that has been carefully hand-optimized for each architecture over many years. Furthermore, the implementation in Impala has one order of magnitude less lines of code than a corresponding hand-tuned expert code. We also obtained similar first results in other domains.

3.11 Design Space Exploration: Getting the Most out of Accelerators

Xiaobo Sharon Hu (University of Notre Dame, US)

For many applications, it is a non-trivial task to actually achieve the high performance and energy efficiency promised by heterogeneous platforms. Introduction of a variety of programmable (such as GPU) or trainable/tunable (such as neural network and constrained optimization solver based) accelerators further exacerbate the problem. One reason is the lack of reliable prediction of the system's performance/energy before application implementation. Another reason is that a heterogeneous platform presents a large design space for workload partitioning among different processing units. Yet another reason is the complicated data usage patterns occurring in many applications.

This talk uses a medical image analysis application as a motivational example to show how different types of accelerators (particularly fully convolutional neural networks and Boolean satisfiability solver) can be employed to solve the problem efficiently and the challenges faced by the design exploration effort. I then present our effort in developing a framework to assist application developers to identify workload partitions that have high potential leading to high performance or energy efficiency for CPU+GPU system *before actual implementation*. The framework can further be used to estimate the performance or energy of given workload partitions. I end the talk with some insights on how such a framework together with our benchmarking approach may be leveraged to help explore the design space of heterogeneous systems with neural network and SAT solver based accelerators.

3.12 Architecture and software for when there's no longer plenty of room at the bottom

Paul H. J. Kelly (Imperial College London, GB)

Joint work of Paul H. J. Kelly, David Ham, Fabio Luporini, Lawrence Mitchell, Mike Giles, Gihan Mudalige, Istvan Reguly, Doru Bercea, Graham Markall, Florian Rathgeber, Maciej Piechotka
Main reference F. Rathgeber et al., "Firedrake: Automating the Finite Element Method by Composing Abstractions", ACM Transactions on Mathematical Software (TOMS), 43(3):24, ACM, 2017.
URL https://doi.org/10.1145/2998441

In 1959, Richard Feynmann wrote his prescient article "Plenty of room at the bottom", demonstrating just how far contemporary computers were from fundamental physical limits. The 58 years of exponential progress since then have brought us much closer to such limits, and there is much debate about where they really lie. What is clear is that we're a lot closer. We are confronted more and more with fundamental physical concerns, particularly with regard to the communication latency, bandwidth and energy. This talk offered a reflection on how this impacts how we think about algorithms and how we design high-performance software. Along the way I discussed the "Turing Tax" – the price we pay for running a programs on a universal, general-purpose machine. I also sketched some of the experience from our lab on delivering software tools that help abstract locality, expose the algorithmic-level design space, and enable tight control over data movement even in code based on irregular data such as unstructured meshes.

3.13 Towards Future-Proof (Parallel?) Programming Models

Christoph W. Kessler (Linköping University, SE)

Joint work of Usman Dastgeer, Johan Enmyren, August Ernstsson, Lu Li, Christoph Kessler
Main reference A. Ernstsson, L. Li, C. Kessler, "SkePU 2: Flexible and type-safe skeleton programming for heterogeneous parallel systems", Int. J. Parallel. Prog., pp. 1–19, Springer, 2017.
URL http://dx.doi.org/10.1007/s10766-017-0490-5

In the coming years, Moore's Law is expected to slow down and current CMOS-based hardware technology to eventually hit physical limits. However, there is, in spite of many exciting new developments, no disruptive replacement technology ready yet to take over after CMOS in a short-term range. Instead, it is likely that the already existing trends towards more parallel/distributed, less coherent, more heterogeneous, less fault-tolerant and more reconfigurable architectures will accelerate further in the coming years. Also, as hardware performance growth declines, an increasingly large share of performance boost must come from improvements in the software, e. g. by more adaptive algorithms and data structures, and more powerful optimizing compiler and runtime system techniques. At the same time, we have to care about portability and programmer productivity to sustain a scalable software market.

In this talk we consider two architecture-independent, high-level (parallel) programming models that can be effectively mapped to today's already quite diverse computer architectures – and hopefully also to coming generations of computing technology:

(1) Skeleton Programming – high-level, customizable general-purpose or domain-specific program building blocks representing frequently occurring patterns of control and data flow, exposing a sequential-looking, compositional programming interface, with adaptive implementations for a broad range of parallel, distributed and heterogeneous target systems. As an example, we briefly review SkePU [1, 2].

(2) Coarse-Grain Dataflow Programming – expressing a computation as a graph of tasks or actors with explicit dependences, which can be configured and mapped statically or dynamically to the resources of a parallel, distributed and heterogeneous target system.

We survey these two complementary approaches, with some techniques in compiler and runtime support for today's architectures, and motivate why they may still be useful with tomorrow's (mostly unknown) computer hardware and can thus lead to more future-proof software.

References

1 C. Kessler et al. *SkePU: Autotunable Multi-Backend Skeleton Programming Framework for Multicore CPU and Multi-GPU Systems.* https://www.ida.liu.se/labs/pelab/skepu/, 2016.
2 A. Ernstsson, L. Li, C. Kessler. *SkePU 2: Flexible and type-safe skeleton programming for heterogeneous parallel systems.* Int. J. Parallel. Prog., 2017.

3.14 Optimization through Hardware and Software Co-Designs

Tei-Wei Kuo (National Taiwan University – Taipei, TW)

This presentation is inspired by the changing of landform by new memory technology, huge driving forces of various applications and heterogeneous computing. There are even more opportunities for system optimization right now than ever. Such optimization opportunities also unsurprisingly exist from the application layer all the way to the system and hardware layers. Excellent examples are software-controlled cache and smart storage devices. In the past decades, we have been experiencing huge impacts due to storage innovation (with a good example on solid-state disks). Some emerging non-volatile memory is now bringing innovation to traditional memory management. Because of that, we soon see the blurring of system boundaries in the memory architecture. With those in mind, challenges and opportunities are coming.

3.15 Future challenges and opportunities for adaptive HPC applications

Matthias Lieber (TU Dresden, DE)

Due to the limits of CMOS technology, high performance computers have experienced a large growth in complexity over the last 15 years: concurrency is exploding (millions of cores in the top systems today), heterogeneous architectures that include accelerators are common, reliability has become a major concern, and performance variability of CPUs can be observed. On the other side, the scientific and engineering applications that run on these systems are also becoming more complex and adaptive, often leading to workload variations over runtime. Dynamic load balancing is used to redistribute dynamic workloads to reduce wasted time in (necessary) synchronization points. Load balancing on such highly parallel computers with heterogeneous compute resources in each node is challenging and trade-offs have to be made regarding workload balance, communication optimization, migration reduction, and the actual costs for making such load balancing decisions. Solutions that focus on some aspects of the problem have been demonstrated, for example very fast load balancing methods can be implemented with space-filling curves and task-based programming models (such as PaRSEC and StarPU) enable load balancing in heterogeneous computing environments. The limits of CMOS technology and possible solutions that have been discussed during this seminar lead to future challenges but also opportunities for high performance computing. Regarding logic, there is not yet a clear successor for CMOS on the horizon. That means that in near future performance gains can only be achieved through architectural improvements and specialization, such as FPGAs and ASICs. However, the increasing heterogeneity will complicate programming as well as runtime workload management. Regarding memory, non-volatile memories will very likely enable improved fault-tolerance mechanisms and accelerate data-intensive applications. 3D stacking of (several layers of memory) on top of logic, as already available in some accelerators, will lead to improved memory performance, potentially reducing the gap between memory and compute performance.

3.16 Neural network-based accelerators: do device, circuits, architectures, or algorithms provide the best "bang for our buck"?

Michael Niemier (University of Notre Dame, US)

Joint work of Michael Niemier, X. Sharon Hu

Researchers are increasingly looking for new ways to exploit the unique characteristics of emerging devices (e. g., non-volatility) as well as architectural paradigms that transcend the energy efficiency wall. In this regard, there has been great interest in hardware implementations of various types of neural networks. One obvious example is IBM's TrueNorth chip, which realizes a configurable, spiking neural network (SNN) in hardware [1]. Additionally, cellular neural networks (CeNNs) are now under investigation via the Semiconductor Research Corporation's benchmarking activities [2, 3] as (i) they can solve a broad set of problems [4] (e. g,. image processing, associative memories, etc.), and (ii) can exploit the unique properties of both spin- and charge- based devices [5, 6]. However, in all cases, we must consider what application spaces/problem sets/computational models ultimately benefit from hardware realizations of neural networks, and if hardware implementations can ultimately outperform alternative architectures/models for the same problem.

This talk will discuss strategies for quantitatively assessing neural network co-processors. To facilitate discussion, as a representative case study, we will consider work with CeNNs. More specifically, we will discuss (i) algorithm development where processing tasks are mapped to CeNNs or more conventional CPUs/GPUs (e. g., for image recognition, CeNNs can be highly efficient for feature extraction tasks given the architecture's parallel nature; for more mathematical operations, CPUs may be more efficient.) (ii) Next, given the analog nature of a CeNN, and the inherent nature of inference applications, algorithmic accuracy must be evaluated at multiple levels (e. g., we must address overall algorithmic quality, and any impact on algorithmic quality due to lower precision hardware.) (iii) Algorithms must then be mapped to a suitable hardware architecture (e. g., parallel CeNNs vs. a CeNN that is used serially). (iv) Finally, we must compare energy, delay, and accuracy projections to the best von Neumann algorithm for the same application-level problem. Using targeting tracking as a case study, we will discuss (a) strategies for algorithmic refinement, and (b) where we derive the most substantial benefits for metrics of interest (energy and delay) – i. e., from devices, circuits, architectures, and/or algorithms.

We will also present preliminary results as to how CeNNs can be leveraged to accelerate convolutional neural networks. As a case study, we present preliminary data for the MNIST digit recognition problem. We will compare/contrast our projections to other architectures and algorithms – e. g., the DropConnect algorithm [7] (with power profiling done on an Intel i5 processor; devices have similar feature sizes to those used in CeNN simulations), IBM True North [8], etc.

References

1 P. A. Merolla et al. *A million spiking-neuron integrated circuit with a scalable communication network and interface.* Science, 345(6197):668–673, 2014.
2 SRC benchmarking center, 2016. https://www.src.org/program/nri/benchmarking/
3 SRC-DARPA STARnet initiative, 2016. https://www.src.org/program/starnet/
4 L. O. Chua and T. Roska. *Cellular Neural Networks and Visual Computing: Foundations and Applications.* Cambridge University Press, 2002.

5 C. Pan and A. Naeemi. *A proposal for energy-efficient cellular neural network based on spintronic devices.* IEEE Transactions on Nanotechnology, 15(5):820–827, 2016.
6 B. Sedighi et al. *A CNN-inspired mixed signal processor based on tunnel transistors.* In DATE 2015.
7 S. Esser et al. *Backpropagation for energy-efficient neuromorphic computing.* In NIPS 2015.
8 L. Wan et al. *Regularization of Neural Networks using DropConnect.* In ICML 2013.

3.17 Scenario-based, System-level Embedded Systems Design

Andy D. Pimentel (University of Amsterdam, NL)

Main reference A. D. Pimentel, "Exploring Exploration: A Tutorial Introduction to Embedded Systems Design Space Exploration", IEEE Design & Test, 34(1):77–90, 2017.
URL http://dx.doi.org/10.1109/MDAT.2016.2626445

Modern embedded systems are becoming increasingly multifunctional and, as a consequence, they more and more have to deal with dynamic application workloads. This dynamism manifests itself in the presence of multiple applications that can simultaneously execute and contend for resources in a single embedded system as well as the dynamic behavior within applications themselves. Such dynamic behavior in application workloads must be taken into account during the design of multiprocessor system-on-a-chip (MPSoC)-based embedded systems. In this talk, I will present the concept of application workload scenarios to capture application dynamism and explain how these scenarios can be used for searching for optimal mappings of a multi-application workload onto an MPSoC. To this end, the talk will address techniques for both design-time mapping exploration as well as run-time mapping of applications.

3.18 Organic electronics: devices for the electronic gadgets age

Sebastian Reineke (TU Dresden, DE)

Organic electronics is a rapidly growing research field that currently develops in the shadow of existing CMOS technology. Ultimately, it targets for applications beyond standard electronics we know today. Success has been achieved already with organic light-emitting diodes (OLEDs), which we find today in most of our mobile displays, and organic solar cells. However, those technologies may soon become commodities, where organic electronics will touch new ground with its potential to deliver scalable, low-cost, customizable (in form and function), and disposable devices. Only future will tell, in which sectors organic electronics will be most successful, but definitely, the route to success proves to be bumpy due to the lack of concerted developments. In this talk, I will give a brief introduction to the conventional approach in organic electronics using the example of OLEDs. Here, I will summarize the general research challenges, give details to some recent device concepts, and assemble a collection of potential scenarios of use.

The main conclusion of this presentation and the discussion related to it is the fact that organic electronics is a very front-end rich technology platform. With attributes like flexible, ultra-lightweight, plastic etc., organic electronics will open up new application scenarios, where the next big challenge is the seamless integration of back-end electronics components made of organic devices that are needed to run such novel front-end applications. Only with the knowledge of the future systems, questions of software development can be addressed in a meaningful fashion. Here, the big question is how diverse the future organic electronics will look like.

3.19 Towards Next Generation of Computing

Heike E. Riel (IBM Research Zurich, CH)

In the past 50 years computing was driven by "smaller & denser" resulting in "faster & cheaper". Cost per function has decreased tremendously, while system performance and reliability have been improved significantly. Dimension scaling alone is no longer sufficient and various paths are pursued in order to increase system performance. In order to further extend core logic and memory technology roadmaps by miniaturization significant innovation in materials, devices and architectures is required. Key technologies which are investigated to continue the roadmap are, e. g., gate-all-around nanowire technologies, III-V semiconducting nanowires for high-mobility field-effect transistors (FETs), III-V nanowires heterostructure tunnel FETs as steep slope devices or carbon nanotube field-effect transistors. In parallel other technologies to build new architectures such as heterogeneous integration, 3D packaging, system-on-chip, silicon photonics and others are pushed to increase system level performance. Yet despite all of these innovative technologies, increasing the density of transistors will cease when length-scales reach atomic dimensions. This raises the fundamental question of what is next? What is the future of information technology beyond scaling and traditional computing? In that regard completely new computing paradigms are developed such as quantum computing and neuromorphic technologies.

3.20 Making better transistors: beyond yet another new materials system

Mark Rodwell (University of California – Santa Barbara, US)

In ~55 years of IC development, industry has concentrated on making switches smaller; universities have mostly concentrated on the complementary role of making them from new materials. Thus has university electron device research progressed from silicon to germanium and SiGe, the arsenide, phosphide and then antimonide III-V's, then carbon nanotubes, and today 2D semiconductors. Beyond SiGe, there seems but little hope that these more recent materials might benefit transistors in computer ICs.

Perhaps we should focus instead on improving their shape? Corrugating a FET channel, in the style of a folded piece of paper, produces a device with transport distance much larger than its footprint; we can use this to improve electrostatics and suppress source-drain tunneling currents in few-nm-footprint transistors. Corrugating the channel in the perpendicular direction increases the drive current per unit IC area, which we then might trade for lower-voltage, lower-power operation. With FETs, making low-resistance yet small contacts is as much a problem as is making short gates: should we corrugate the metal-semiconductor interface to reduce the interface resistance?

Or, should we change their band structure? In tunnel FETs, we can add several heterojunctions, and so increase the desired on-state tunneling currents while decreasing the unwanted off-state leakage currents. Yet, I can offer nothing beyond this single example, albeit one that I presently find of great interest.

Finally, perhaps we might change their function? One focus of this workshop is to explore the merging of logic and memory. We might do this at either within the transistor or within low-level logic design. It is not yet clear to this circuit designer that a logic-plus-memory transistor would be markedly more useful than, for example, simple merged logic using pass transistors and gate capacitances. I will do my best to examine their potential utility.

3.21 Logic Synthesis for Post-CMOS Technologies

Eleonora Testa (EPFL – Lausanne, CH), Giovanni De Micheli (EPFL – Lausanne, CH), and Mathias Soeken (EPFL – Lausanne, CH)

Traditionally, logic synthesis tools have concentrated on the optimisation of circuits based on CMOS logic primitives such as AND and OR. Since recently many emerging nanotechnologies are based on logic models different from standard CMOS, new logic synthesis approaches need to be considered. Most of the promising nanodevices, such as Resistive Random Access Memories (RRAMs) and Spin Wave Devices (SWDs), are based on majority logic and are characterised by nontrivial technological constraints. Both aspects are fundamental when designing new logic synthesis tools. In this talk, we present how many emerging technologies can benefit from a majority-based logic synthesis approach [1, 2]. We will concentrate on a new data structure that provides the necessary abstraction for Boolean functions optimization and manipulation [3]. Further, we will illustrate how SAT-based methods can be used to address the technological constraints.

References

1 E. Testa et al. *Inversion optimization in majority-inverter graphs.* In NANOARCH 2016.
2 O. Zografos et al. *Design and benchmarking of hybrid CMOS-spin wave device circuits compared to 10nm CMOS.* In IEEE-NANO 2015.
3 L. Amarù et al. *Majority-inverter graph: a novel data-structure and algorithms for efficient logic optimisation.* In DAC 2014.

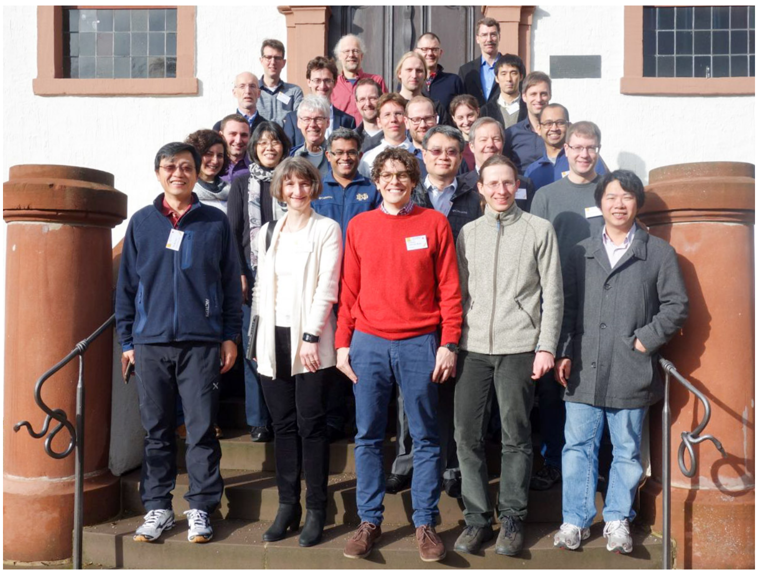
3.22 Architectural Requirements for Intransitive Trust and Fault and Intrusion Tolerance

Marcus Völp (University of Luxembourg, LU)

Looking at our software stack today, we are facing a functionality / code size dilemma: functionality comes with a certain amount of code and the more code, the more vulnerabilities and possibilities for attackers to compromise the computer systems we all depend on. In this talk, I review designs for intransitive trust relationships, which allow critical applications to use functionality without trusting all the code that provides this functionality. Common patterns for intransitive trust involve ciphers to protect data integrity and confidentiality. Another involves replication and voting to hide Byzantine behavior of a minority of compromised replicas behind consensus of a healthy majority. I derive architectural implications and raise as questions how the strong isolation assumptions of intransitive trust design patterns can be realized in todays and upcoming wildly-heterogeneous systems.

Participants

- Rehan Ahmed
ETH Zürich, CH
- Stefano Ambrogio
IBM Almaden Center – San Jose, US
- Nils Asmussen
TU Dresden, DE
- Tal Ben-Nun
The Hebrew University of Jerusalem, IL
- Jerónimo Castrillón-Mazo
TU Dresden, DE
- Jian-Jia Chen
TU Dortmund, DE
- Yiran Chen
Duke University – Durham, US
- Martin Claus
TU Dresden, DE
- Ayse Coskun
Boston University, US
- Suman Datta
University of Notre Dame, US
- Erik P. DeBenedictis
Sandia National Labs – Albuquerque, US
- Gianluca Fiori
University of Pisa, IT
- Shunsuke Fukami
Tohoku University, JP
- Balazs Gerofi
RIKEN – Tokyo, JP
- Sebastian Hack
Universität des Saarlandes, DE
- Xiaobo Sharon Hu
University of Notre Dame, US
- Paul H. J. Kelly
Imperial College London, GB
- Christoph W. Kessler
Linköping University, SE
- Tei-Wei Kuo
National Taiwan University – Taipei, TW
- Roland Leißa
Universität des Saarlandes, DE
- Max Lemme
Universität Siegen, DE
- Matthias Lieber
TU Dresden, DE
- Michael Niemier
University of Notre Dame, US
- Stuart Parkin
MPI – Halle, DE
- Andy D. Pimentel
University of Amsterdam, NL
- Sebastian Reineke
TU Dresden, DE
- Heike E. Riel
IBM Research Zurich, CH
- Mark Rodwell
University of California – Santa Barbara, US
- Eleonora Testa
EPFL – Lausanne, CH
- Marcus Völp
University of Luxembourg, LU



Report from Dagstuhl Seminar 17062

Beyond VR and AR: Reimagining Experience Sharing and Skill Transfer Towards an Internet of Abilities

Edited by

Anind K. Dey[1], Jonna Häkkilä[2], Niels Henze[3], and Kai Kunze[4]

1 **Carnegie Mellon University, US** anind@cs.cmu.edu
2 **University of Lapland, FI** jonna.hakkila@ulapland.fi
3 **Universität Stuttgart, DE** niels.henze@vis.uni-stuttgart.de
4 **Keio University, JP** kai@kmd.keio.ac.jp

Abstract

With recent development in capture technology, preserving one's daily experiences and one's knowledge becomes richer and more comprehensive. Furthermore, new recording technologies beyond simple audio/video recordings become available: 360° videos, tactile recorders and even odor recorders are becoming available. The new recording technology and the massive amounts of data require new means for selecting, displaying and sharing experiences. This seminar brought together researchers from a wide range of computing disciplines, including virtual reality, mobile computing, privacy and security, social computing and ethnography, usability, and systems research. Through lightning talk, thematic sessions and hands-on workshops, the seminar investigated the future of interaction beyond virtual and augmented reality. Participants reimagined experience sharing and skill transfer towards an Internet of abilities. We conclude with a set of open and guiding questions for the future of our field.

Seminar February 05–10, 2017 – http://www.dagstuhl.de/17062
1998 ACM Subject Classification H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous
Keywords and phrases Augmented Reality, Virtual Reality, Skill Transfer
Digital Object Identifier 10.4230/DagRep.7.2.23
Edited in cooperation with Yun Suen Pai

1 Executive Summary

Anind K. Dey
Jonna Häkkilä
Niels Henze
Kai Kunze

Sharing experiences and knowledge have always been essential for human development. They enable skill transfers and empathy. Over history, mankind developed from oral traditions to cultures of writing. With the ongoing digital revolution, the hurdles to share knowledge and experiences vanish. Already today it is, for example, technically feasible to take and store 24/7 video recordings of one's life. While this example creates massive collections of data, it makes it even more challenging to share experiences and knowledge with others in meaningful ways. Facilitating the third wave of VR and AR technologies we are currently witnessing, researchers started to broadly look at VR and AR again. Topics ranging from using AR to mitigate skills gaps [1] and understanding user interaction with commercial AR games [4], to using focus depth as an input modality for VR [2], and understanding the effect of gender in VR [3]. The goal of the seminar was to take a step back from the technical research to look at the fundamental aspects of interactive media.

A recurring theme in science fiction literature is the act of downloading another human's abilities to one's mind. Although current cognitive science and neuroscience strongly suggest that this is impossible, as our minds are embodied; we believe that skill transfer and effective learning will accelerate tremendously given recent technological trends; just to name a few of the enabling technologies, human augmentation using virtual/augmented reality, new sensing modalities (e.g. affective computing) and actuation (e.g. haptics), advances in immersive storytelling (increasing empathy, immersion, communication) etc.

Ultimately, we believe this will lead to "downloadable" experiences and abilities. The effects will definitely not be instant and it will most likely be very different from the Sci-Fi theme. Yet, these differences are exactly what we want to explore in this seminar. Computer scientists in wearable computing, ubiquitous computing, human computer interaction, affective computing, virtual reality and augmented reality have been working on related topics and enabling technologies for years. However, these developments are disjointed from each other. With this seminar we want to bring them together working in the virtual/augmented/mixed reality, ubiquitous computing, sensing and HCI fields discussing also with experts in cognitive science, psychology and education.

While sharing experiences and knowledge through communication and socializing are a long time focus of various research efforts, we believe it is necessary to rethink and redefine experience sharing and skill transfer in light of the following current technological advances like the following:

1. Affordable Virtual Reality and Augmented Reality systems will become available to consumers in the near future (or already are available).
2. Advances in new sense sharing technologies (e.g. eye gaze, haptics, odors).
3. Advances in real-life tracking of physical and cognitive activities and emotional states.
4. Educators, cognitive scientists and psychologists have now a better understanding of individual and group behaviors, empathy and fundamentals of learning.

The seminar was structured around lightning talks by the participants, two hands-on workshops and three thematic sessions. In the lightning talks, the participants introduced themselves and shared their vision with the group. The first hands-on workshop by Shunichi Kasahara introduced the term Superception and showcased prototypes in this domain. The second workshop organized by Pedro Lopes enabled participants to experiment with electrical muscle stimulation by connecting off-the-shelf devices to embedded systems. Three days of the seminar started with thematic sessions run by one of the organizers. The sessions explored the future of human-computer symbiosis, human augmentation, and enabling technologies.

References

1 Yamashita, S., Matsuda, A., Hamanishi, N., Suwa, S., & Rekimoto, J. (2017, March). Demulti Display: A Multiplayer Gaming Environment for Mitigating the Skills Gap. In Proceedings of the Tenth International Conference on Tangible, Embedded, and Embodied Interaction (pp. 457–463). ACM.

2 Pai, Y. S., Outram, B., Vontin, N., & Kunze, K. (2016, October). Transparent Reality: Using Eye Gaze Focus Depth as Interaction Modality. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (pp. 171–172). ACM.

3 Schwind, Valentin; Knierim, Pascal; Tasci, Cagri; Franczak, Patrick; Haas, Nico; Henze, Niels. "These are not my hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality Inproceedings CHI '17 Proceedings of the 2017 Annual Symposium on Computer-Human Interaction, ACM Press, New York, NY, USA, 2017, ISBN: 9781450346559,

4 Colley, Ashley, Jacob Thebault-Spieker, Allen Yilun Lin, Donald Degraen, Benjamin Fischman, Jonna Häkkilä, Kate Kuehl, Valentina Nisi, Nuno Jardim Nunes, Nina Wenig, Dirk Wenig, Brent Hecht and Johannes Schöning. 2017. The Geography of Pokémon GO: Beneficial and Problematic Effects on Places and Movement. In Proc. CHI'17.

2 Table of Contents

3 Overview of Talks

3.1 Towards Unremarkable Use of Augmented and Virtual Reality

Ashley Colley (University of Lapland, FI, Ashley.Colley@ulapland.fi)

Augmented Reality (AR) is currently taking its first real steps out from controlled laboratory environments into the wild. The global PokemonGo phenomena in the summer of 2016 raised questions related to the definition of AR. Rather than focusing on visual aspects, as highlighted by Azuma's definition of AR, e.g. "...registered in 3D", for many users the perception of the real world overlaid with a layer of virtual content was the dominant perception. Study of PokemonGo revealed that location based advantages in the real world were transferred to the virtual content [1]. The smartphone has become the current de facto method of AR browsing in the wild, based on the suitability of their features such as high resolution camera, GPS and inertial sensors. However, smartphones were not designed with the AR browsing task in mind, and this results in a less than optimal user experience [2]. One approach when designing the optimal handheld AR browser device is to examine the balance between the goal of Azuma's perfectly aligned virtual and real content, and practical and ergonomic considerations for in the wild usage. By creating a handheld AR browser device where the device's camera is at a 45 degree angle to its display, efficient AR browsing can be achieved without the physical load of holding the device at eye-level as a magic-lens [3]. The next steps in AR should aim to address in-the-wild usage, focusing on the overall user experience, including multi-sensory and social aspects, rather than perfecting the visual experience. At the same time the potential of the virtual world to provide a more equal experience than the real world should be a core tenet directing the evolution of the domain.

References

1 Colley, Ashley, Jacob Thebault-Spieker, Allen Yilun Lin, Donald Degraen, Benjamin Fischman, Jonna Häkkilä, Kate Kuehl, Valentina Nisi, Nuno Jardim Nunes, Nina Wenig, Dirk Wenig, Brent Hecht and Johannes Schöning. 2017. The Geography of Pokémon GO: Beneficial and Problematic Effects on Places and Movement. In Proc. CHI'17.
2 Ashley Colley, Tuomas Lappalainen, Elisa Määttänen, Johannes Schöning, and Jonna Häkkilä. 2016. Crouch, Hold and Engage: Spatial Aspects of Augmented Reality Browsing. In Proc NordiCHI '16
3 Ashley Colley, Wouter Van Vlaenderen, Johannes Schöning, and Jonna Häkkilä. 2016. Changing the camera-to-screen angle to improve AR browser usage. In Proc MobileHCI'16.

3.2 Augmented Reality for Sensemaking

Nitesh Goyal (Cornell University – New York, US, ngoyal@cs.cornell.edu)

Complex world problems of the future will be solved by encouraging collaboration between humans. However, the present VR/AR frameworks lack seamless collaboration frameworks for problem-solving or decision-making between collaborators. Future AR systems will need to be merged into the collaboration technologies' setup to enable such complex tasks without significant cognitive load, yet enhancing task performance. Sharing has been shown to be tricky, especially when required to do so explicitly between collaborators. Implicit sharing has been shown to improve task performance. Future AR systems can benefit from implicit sharing of information synchronously or asynchronously. This would enable collaborators to leverage peripherally shared information. Alternatively, AR systems may play not just a passive role, but an active role too. Identifying and sharing relevant information, embedded in the real world actively to encourage awareness is what I anticipate the future of AR to look like. The challenges that prevent us from reaching this goal, include technical challenges like the lack of an AR equipped environment, but also include socio-technical challenges that assume that future users will be expert at using AR systems, and subsequently will use the systems ethically.

3.3 Involving Users in Future Visions

Jonna Häkkilä (University of Lapland, FI, Jonna.Hakkila@ulapland.fi)

We live in a world where the technological innovations are advancing in a rapid speed. Mobile technology has fundamentally changed our everyday life during the past two decades, and we are now able to access other people, information and different types of services whenever we wish. Now the smart phone is the primary general ICT tool and user interface (UI) while mobile, but next steps are already emerging to the use for large audiences. Wearable technologies have already been adopted by masses of people e.g. in the form of activity trackers, and form factors such as smart watches and bracelets are a commodity. AR and VR technologies have also become better affordable for developers and consumers e.g. for gaming, and although not yet visible on the streets, products in this frontier are emerging.

When developing novel technology solutions that are aimed for large user groups, it is important to pay attention to the usability and user experience (UX) with the devices and applications. Ease of use, ergonomics, and aesthetic design are factors, which affect to the user's interest and engagement with the technology. The social acceptability should not be neglected, as privacy concerns and embarrassment when using unusual gadgets in public can greatly affect to the adoption of new technologies. The focus on ubiquitous computing research has so far been heavily on the technology side instead of user experience [4]. While technology has become more mature and miniaturized, the possibilities to explore different design aspects have grown. My research addresses the user experience design and user centric design of future technologies. Through design, we can communicate technology visions to large audiences and create concepts, which appeal to people as potential future garments or products, not just as engineering demos. An example of such is Solar Shirt design concept and prototype, an environmental awareness wearable utilizing printed electronics solar cells and flexible displays are part of the design of a fashion garment [3]. User experience design can also seek novel materials for interaction. In BreathScreen concept and prototype shows how a situated fog screen is created from the breath or smoke around the user, forming an ephemeral UI [1]. This kind of novel interfaces can be used to augment us and our immediate surroundings in a pleasant and experience rich manner. In my research I also wish to highlight the importance of evaluating novel technologies in-the-wild with users, as e.g. in our experiment of trying out skiing and snowboarding in VR in-the-wild, i.e. in a downhill slope [2]. By exposing the concepts to a real life use context, we gain valuable insights of its requirements and challenges.

References

1 Ismo Alakärppä, Elisa Jaakkola, Ashley Colley, Jonna Häkkilä. 2017. BreathScreen – Design and Evaluation of an Ephemeral UI. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). ACM, New York, NY, USA.

2 Ashley Colley, Jani Väyrynen, Jonna Häkkilä. 2015. Skiing in a blended virtuality: an in-the-wild experiment. In Proceedings of the 19th International Academic Mindtrek Conference (AcademicMindTrek '15), 89–91. ACM.

3 Paula Roinesalo, Lasse Virtanen, Tuomas Lappalainen, Anu Kylmänen, and Jonna Häkkilä. 2016. Solar shirt: design of an environmental awareness wearable. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16). ACM, New York, NY, USA, 495–499. DOI: http://dx.doi.org/10.1145/2968219.2971350

4 Kaisa Väänänen-Vainio-Mattila, Thomas Olsson, Jonna Häkkilä. 2015. Towards Deeper Understanding of User Experience with Ubiquitous Computing Systems: Systematic Literature Review and Design Framework. In Proc. INTERACT 2015. Springer 2015.

3.4 When Information is not Scarce

Niels Henze (University of Stuttgart, DE, niels.henze@vis.uni-stuttgart.de)

We currently witness another wave of augmented- and virtual reality research. For the first time, augmented- and virtual reality technologies become not only widely available to consumers but also adopted by them. In the previous weave, one of our research focus was highlighting content that is currently not in the user's field of view by developing and evaluating off-screen visualisations [1, 2]. In the last years, we not only witness improvements in augmented- and virtual reality technologies, but also face a dramatic change of how humans and computers interact. The classic human-computer interaction principle was based on the assumption that users should start with an action and the computer responses with a reaction. We currently see more and more systems that violate this principle. What started with simple notifications about incoming emails developed into a whole notification ecosystem (see e.g. [3, 4]).

The question today is not how to highlight that more information is available but to support users coping with proactive computing and a large amount of available information. We investigate different directions to support users. In virtual reality, we investigate how virtual representations of the user should be presented to be accepted by users [5, 6]. Regarding notifications, we study how users interact with notifications that are proactively delivered and developed new means to present and manage them [7, 8, 9]. We further investigate approaches that take the user's attention into account when presenting information [10]. Ultimately, it will be necessary to bring the components together and build systems that actively manage the user's attention. Just as computer science developed algorithms to manage other scarce resources such as processing poser and random access memory, we need to develop algorithms that manage human attention.

References

1 Henze, N., & Boll, S. (2010, September). Evaluation of an off-screen visualization for magic lens and dynamic peephole interfaces. In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (pp. 191–194). ACM.

2 Schinke, T., Henze, N., & Boll, S. (2010, September). Visualization of off-screen objects in mobile augmented reality. In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (pp. 313–316). ACM.
3 Sahami Shirazi, A., Henze, N., Dingler, T., Pielot, M., Weber, D., & Schmidt, A. (2014, April). Large-scale assessment of mobile notifications. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 3055–3064). ACM.
4 Shirazi, A. S., & Henze, N. (2015, August). Assessment of Notifications on Smartwatches. In Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (pp. 1111–1116). ACM.
5 Schwind, Valentin; Knierim, Pascal; Tasci, Cagri; Franczak, Patrick; Haas, Nico; Henze, Niels. "These are not my hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality Inproceedings CHI '17 Proceedings of the 2017 Annual Symposium on Computer-Human Interaction, ACM Press, New York, NY, USA, 2017, ISBN: 9781450346559,
6 Schwind, V., Wolf, K., Henze, N., & Korn, O. (2015, October). Determining the characteristics of preferred virtual faces using an avatar generator. In Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (pp. 221–230). ACM.
7 Weber, D., Voit, A., Kratzer, P., & Henze, N. (2016, September). In-situ investigation of notifications in multi-device environments. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 1259–1264). ACM.
8 Voit, A., Machulla, T., Weber, D., Schwind, V., Schneegass, S., & Henze, N. (2016, September). Exploring notifications in smart home environments. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (pp. 942–947). ACM.
9 Weber, D., Shirazi, A. S., & Henze, N. (2015, August). Towards Smart Notifications Using Research in the Large. In Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (pp. 1117–1122). ACM.
10 Dingler, T., Rzayev, R., Schwind, V., & Henze, N. (2016, September). RSVP on the go: implicit reading support on smart watches through eye tracking. In Proceedings of the 2016 ACM International Symposium on Wearable Computers (pp. 116–119). ACM.

3.5 Superception

Shunichi Kasahara (Sony CSL – Tokyo, JP, kasahara@csl.sony.co.jp)

How technologies empower us? How technologies change the human perception? I am right now envisioning a research concept called Super + perception = superception. The word "super" has two meanings : one is to augment, enhance or empower; the other is connection beyond individuals. I am exploring ways to augment and transform our perception by intervening our sensation computationally or connecting multi human perception with technologies. From this point of view, beyond VR and AR means sensory augmentation and substitution beyond visual related technologies. In a broader sense, emergence of sensory related technologies enable us to access the internal data of our body i.e. human perception. With these technologies, we will be able to engineer our perception.

In my research, I explore ways to produce Superception according to three strategies: reproducing perception, using illusionary perception computationally and connect multiple perception and sensation. In this workshop, I presented Parallel Eyes, which is a system that connect multiple first person view so that four person can see the shared first person view as well as their own.

Engineering for accessing digital data using computer and performing real world functions will be interpreted as "Engineering for interface" which includes AR interface. On the other hand, the approach of Superception to control human perception and sensation using sensing technology and virtual reality can precisely control human inputs and outputs, termed "Engineering for Perception". I believe that Superception will be a platform for engineering control related to human senses and augmentation of human abilities and perception.

References

1 Shunichi Kasahara, Mitsuhito Ando, Kiyoshi Suganuma, and Jun Rekimoto. 2016. Parallel Eyes: Exploring Human Capability and Behaviors with Paralleled First Person View Sharing. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 1561–1572. DOI: https://doi.org/10.1145/2858036.2858495

3.6 AR-ready environments

Gudrun Klinker (Technical University of Munich, DE, klinker@in.tum.de)

Up to very recently, research goals of VR and AR have been rather technology-driven: we need to build machines and devices (displays, sensors) that were fast and precise enough to convey the 3D illusion of perceiving virtual objects in 3D environments. Newest technology (HMDs, optical and time-of-flight sensors) is on the verge of passing fundamental thresholds related to human sensing limitations. Going beyond these threshold means creating user experiences suitable for real applications. Research transforms from considering technical aspects towards considering a human perspective: we are surrounded by masses of virtual information. How can computers help us perceive and interact with this information?

At the Technical University Munich, we are developing a framework for Ubiquitous Augmented Reality, which provides users with AR-services wherever they go via ubiquitous tracking, ubiquitous presentation and ubiquitous interaction. This lays the foundation towards creating, evolving and testing many different approaches to have users experience augmented worlds. Users can interact with information based on technology provided by complex hybrid combinations of mobile and stationary devices installed in an AR-ready world.

3.7 Towards Engaging Augmented Reality Environments

Stephan Lukosch (TU Delft, NL, s.g.lukosch@tudelft.nl)

Orson Scott Card impressively shows the effect of high fidelity in simulation games in his novel Ender's Game [3]. The main actor Ender and his team believe they are playing a training simulation game for fighting a war on an alien race. In the end, it becomes clear that Ender was in fact commanding the real fleet through the game, attacking and finally extinguishing the alien race. Tad Williams describes in his saga Otherland [20, 21, 23, ?] a future world with a widespread availability of full-immersion virtual reality [11] installations. These installations allow people to access an online world, called simply 'the Net'. Within the Net, a group of people aims to achieve immortality. In his novel Rainbows End [19] Vernor Vinge describes how the main character Robert Gu is slowly recovering from Alzheimer's disease due to medical advances in the future. While recovering, former technophobe Robert adapts to a changed world in which almost every object is networked and the use of augmented reality [1, 2] is normal. Humans interact within augmented reality by wearing smart clothes and contact lenses that can overlay the physical environment with computer graphics. In Rainbows End [19], augmented reality is used for various purposes, e.g., large-scale commercial gaming areas, supporting maintenance workers with blueprints of machines or buildings, communication with virtual avatars and diagnostic purposes in medical settings.

Science Fiction authors Orson Scott Card, Tad Williams and Vernor Vinge forecast a vision of engaging augmented and virtual reality environments that current research is already addressing. Feng et al. study the effect of wind and vibrations on orientation in virtual environments [8]. Narumi et al. consider the effect of artificial smell and augmented reality on taste [12]. There has been quite some research on introducing smell into movie theaters and television [9] and even more research on haptic feedback [17]. One of the most difficult aspects to reproduce, however, is a realistic interaction with other (real or virtual) humans. Olson and Olson [13, 14] analysed technology support for virtual co-location. They came to the conclusion that distance matters and that the analysed technology is not mature enough to enable virtual co-location. Olson and Olson stated that even future technology will struggle to enable virtual co-location. In their opinion, providing awareness among co-workers and enabling co-reference as well as spatial referencing will remain a challenge. Complex problem solving still requires a team of experts to physically meet and interact with each other. Then, the identification of the problem and the creation of a shared understanding are major challenges for efficiently solving a problem [15]. Typical scenarios are, e.g., solving complex construction problems, training the usage of complex machinery, analysing complex situations in emergency services or diagnosing complex medical situations. Unfortunately, it is not always possible to bring a team together to handle a complex situation. This is due to experts' availability, critical timing issues or accessibility of a location. While in the novel Rainbows End [19], such situations are supported with augmented reality technology, current technology is not yet there.

We have taken first steps towards the combined visions of Orson Scott Card, Tad Williams and Vernor Vinge on highly engaging augmented reality environments. We showed [16, 5, 6] that virtual co-location can enable experts at a distance to interact with investigators on a crime scene and jointly perform investigation tasks. We further showed that such interaction as well as the exchange of information in augmented reality increases the situational awareness of teams [10]. With a game on jointly building a tower out of virtual blocks [7], we showed that virtual co-location can be used to collaboratively solve complex spatial problems. Further, we have combined serious games in augmented reality with sensors for motion tracking [4, 18] to create novel and engaging approaches for human motor function assessment. In future research, we will explore how to address all human senses as described by Tad Williams. Here, we are especially interested in the effect on engagement and the training outcomes when using serious games in augmented reality for training complex scenarios or learning complex tasks. We will continue our research on how to enable realistic interaction between local users as well as remote users in augmented environments and on how to enable interaction between real and virtual objects. Addressing these research issues will allow us to go beyond current VR and AR environments and create engaging augmented reality environments for future experience sharing and skill transfer.

References

1 Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. (2001). Recent advances in augmented reality. Computer Graphics and Applications, IEEE, 21(6):34–47.
2 Azuma, R. T. (1997). A survey of augmented reality. In Presence: Teleoperators and Virtual Environments 6, pages 355–385.
3 Card, O. S. (1985). Ender's Game. New York City, USA: Tor Books.
4 Cidota, M. A., Lukosch, S. G., Dezentje, P., Bank, P. J., Lukosch, H. K., & Clifford, R. M. (2016). Serious gaming in augmented reality using hmds for assessment of upper extremity motor dysfunctions. i-com -Journal of Interactive Media, Special Issue on Smartglass Technologies, Applications and Experiences, 15(2), 155–169.
5 Datcu, D., Cidota, M., Lukosch, H., and Lukosch, S. (2014a). On the usability of augmented reality for information exchange in teams from the security domain. In IEEE Joint Intelligence and Security Informatics Conference (JISIC), pages 160–167.
6 Datcu, D., Cidota, M., Lukosch, H., and Lukosch, S. (2014b). Using augmented reality to support information exchange of teams in the security domain. In IEEE International Symposium on Mixed and Augmented Reality (ISMAR).
7 Datcu, D., Lukosch, S. and Lukosch, H. (2016. A Collaborative Game to Study Presence and Situational Awareness in a Physical and an Augmented Reality Environment. In Journal of Universal Computer Science, 2016, 22, 247–270
8 Feng, M., Lindeman, R. W., Abdel-Moati, H., & Lindeman, J. C. (2015). Haptic ChairIO: A system to study the effect of wind and floor vibration feedback on spatial orientation in VEs. In IEEE Symposium on 3D User Interfaces, 3DUI 2015, Arles, France, March 23-24, 2015 (pp. 149–150).
9 Kim, H., Park, J., Noh, K., Gardner, C. J., Kong, S. D., Kim, J., and Jin, S. (2011). An x-y addressable matrix odor-releasing system using an on-off switchable device. Angewandte Chemie, 123(30):6903–6907.
10 Lukosch, S., Lukosch, H., Datcu, D., & Cidota, M. (2015). Providing information on the spot: Using augmented reality for situational awareness in the security domain. Computer Supported Cooperative Work (CSCW) – The Journal of Collaborative Computing and Work Practices, 24(6), 613–664.
11 Milgram, P. and Kishino, F. (1994). A taxonomy of mixed reality visual displays. IEICE Transactions on Information Systems, E77-D(12).
12 Narumi, T., Kajinami, T., Tanikawa, T., & Hirose, M. (2010). Meta cookie. In ACM SIGGRAPH 2010 Emerging Technologies, SIGGRAPH '10 (pp. 18:1–18:1). New York, NY, USA: ACM Press.
13 Olson, G. M. and Olson, J. S. (2000). Distance matters. Human-Computer Interaction, 15(2-3):139–178.
14 Olson, J. S. and Olson, G. M. (2014). How to make distance work work. interactions, 21(2):28–35.
15 Piirainen, K., Kolfschoten, G., and Lukosch, S. (2012). The joint struggle of complex engineering: A study of the challenges of collaborative design. International Journal of Information Technology & Decision Making (IJITDM), 11(6):1–39.
16 Poelman, R., Akman, O., Lukosch, S., and Jonker, P. (2012). As if being there: Mediated reality for crime scene investigation. In CSCW '12: Proceedings of the 2012 ACM conference on Computer Supported Cooperative Work, pages 1267–1276. ACM New York, NY, USA.

17 Samur, E. (2012). State of the art. In Performance Metrics for Haptic Interfaces, Springer Series on Touch and Haptic Systems, pages 9–26. Springer London.
18 van der Meulen, E., Cidota, M. A., Lukosch, S. G., Bank, P. J., van der Helm, A. J., & Visch, V. (2016). A haptic serious augmented reality game for motor assessment of Parkinson's disease patients. In IEEE International Symposium on Mixed and Augmented Reality Adjunct Proceedings (pp. 102–104). Washington, DC, USA: IEEE Computer Society.
19 Vinge, V. (2007). Rainbows End. Tor Books, New York City, USA.
20 Williams, T. (1996). Otherland – City of Golden Shadow. Legend Books, London, UK.
21 Williams, T. (1998). Otherland – River of Blue Fire. Legend Books, London, UK.
22 Williams, T. (1999). Otherland – Mountain of Black Glass. Legend Books, London, UK.
23 Williams, T. (2001). Otherland – Sea of Silver Light. Legend Books, London, UK.

3.8 Cognitive Engineering for VR and AR applications

Joachim Meyer (Tel Aviv University, Israel, jmeyer@tau.ac.il)

Much of the research on AR and VR focuses on technological development, but increasingly it becomes possible to deploy solutions for actual use. These possibilities raise the question how one should design VR and AR to fit into specific systems and contexts. The approaches developed in cognitive engineering and human performance modeling can help us determine the optimal design of systems for a given context. It is possible to address many problems with these modeling methods, such as:

1. What is the required level of reality? For some purposes, a very crude, low-resolution imagery may be sufficient, while in other contexts, an application can only function if it responds extremely fast and has very high graphic quality. For instance, in research on driving one can use low-resolution PC-based displays of the roadway to study the effects of in-vehicle devices on driver distraction (e.g., [1]). In contrast, to study driver responses to vehicle dynamics, one needs very advanced moving-base simulators, or one may actually need real vehicles.
2. What are the required and what are optimal properties of information displays (such as alerts) in AR? The design of these systems needs to balance the possibility for alarm fatigue (the cry-wolf syndrome) resulting from excessively frequent alerts, as opposed to the possibility of complacency, where people rely very strongly on alerts and fail to monitor other information [2].
3. What settings of the system can users adjust correctly and what settings should be determined for the user? Often users may not have the necessary information to decide on the correct setting (such as a threshold) [3]. The use of models for these (and related) issues can help in all stages of the system life cycle, from the initial specifications, over the design, the deployment, up to the evaluation of the functioning system.

References

1 Lavie, T., & Meyer, J. (2010). Benefits and costs of adaptive user interfaces. International Journal of Human-Computer Studies, 68, 508–524.
2 Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. Human Factors, 46, 196–204.
3 Meyer, J., & Sheridan, T. B. (in press). The intricacies of user adjustments of alerting thresholds. Human Factors.

3.9 Digital Humanities

Yannick Prié (University of Nantes, FR, yannick.prie@univ-nantes.fr)

I am a professor of computer science at the University of Nantes in France. I am currently involved in the installation of the University into a new building dedicated to research, teaching and innovation around Digital Culture, with interdisciplinarity at its core. There will also be an associated masters program in relation with schools of Art, Design, and Communication. I began working on ontological and document engineering and I moved towards HCI and data visualization, and my motto has now evolved to "computer science & interdisciplinary". I have been working in the field of digital humanities, with current projects in science and technology studies, and in learning analytics [1]. I have also worked in trace technology, doing video active reading and annotation systems such as the open source video annotation software Advene [2] as well as trace-based reflective systems [3, 4]. I have lately turned toward visual analytics, and have several project related to interactive and progressive mining [5], as well as VR-based immersive analytics. I have also been interested into studying activity development [6]; both from a 3rd person and a 1st person perspective (micro-phenomenology), and have recently began working with psychotherapists around tdcs and VR.

References

1 Madjid Sadallah, Benoît Encelle, Azze-Eddine Maredj, Yannick Prié. Towards Reading Session-based indicators in Educational Reading Analytics. in 10th European Conference on Technology Enhanced Learning, Toledo, Spain, sept. 2015, 14p.
2 Olivier Aubert and Yannick Prié (2005) Advene: Active Reading through Hypervideos. in ACM Conference on Hypertext and Hypermedia 05, 2005, pp. 235–244.
3 Elise Lavoué, Gaëlle Molinari, Yannick Prié, Safè Khezami. Reflection-in-Action Markers for Reflection-on-Action in Computer-Supported Collaborative Learning Settings. Computers & Education. Volume 88, October 2015, pp. 129–142.
4 Leila Yahiaoui, Yannick Prié, Pierre-Antoine Champin, Zizette Boufaïda. Redocumenting computer mediated activity from its traces: a model-based approach for narrative construction. in Journal of Digital Information, vol. 12. n°3. 2011.
5 Vincent Raveneau, Julien Blanchard, Yannick Prié. Pattern-based progressive analytics on interaction traces. Poster at IEEE VIS 2016, Baltimore. oct. 2016. 2p.
6 Amaury Belin, Yannick Prié. Towards a model for describing appropriation processes through the evolution of digital artifacts. in Designing Interactive Systems DIS2012, pp. 645–654, Newcastle, June 2012.

3.10 Physiological Signal-Driven Virtual Reality in Social Spaces

Pai Yun Suen (Keio University, JP, yspai1412@gmail.com)

VR and AR has been progressing at a rapid pace, and even though current mechanics are not perfect, such as locomotion [1] or haptic systems, its continuous evolution requires further design considerations for new interaction mechanics. One of the key issues to consider is for a future where VR and AR are being used anywhere, anytime. Foreseeing such a future in the next 10 years is a very likely outcome, and new issues will arise from such a usage. For instance, the interaction that is often accustomed for VR; reaching out to grab and object, performing wide gestures, etc. These interaction methods are immersive, but highly unlikely to be usable in public areas. Prior to gestures, buttons on controllers were the traditional input mechanic, but this on the other hand suffers from realism and to a higher degree, a cause for simulation sickness.

To accommodate these social spaces and proxemics, the idea of subtle input and subtle interactions are a worthwhile investigation. This refers to inputs and interactions that are unobtrusive and won't annoy others. Firstly, it is important to determine the definition of a social space. What are the physical constrains that we face in our daily life when navigating in the physical environment? In proxemics, human space can be categorized under 4 types; intimate, personal, social and public. Furthermore, depending on the scenario, these spaces have further constrains. For instance, standing in the bus requires us to be holding a handle for safety, meaning interactions are preferably hands-free. The second issue is concerning the preferred input methods by the users. Physiological sensing [2] is an interesting form of input because it provides both explicit and implicit data regarding our current physiological state, to be used as a mechanic in VR or AR. Input methods such as using eye gaze [3, 4], muscle contraction, or even brain interface are worth investigation to determine the users' preferences. The third issue then would be to determine the appropriate interaction mechanic for these sensing methods to remain subtle and unobtrusive. For example, eye gaze is suitable for selection in a virtual environment, but less preferable for activation to avoid straining the eye.

The concept of social acceptance toward VR being "anytime, anywhere" actually encompasses several other considerations that are also worth mentioning, such as how the difference in culture may impact the kind of sensing methods that are allowed to be used in public spaces. Furthermore, a solution is required for interactions in the virtual word to simply "blend" into real world interactions. A simple example would be if we were required to tie a shoe lace while wearing a HMD. It would be a hassle to remove the HMD to tie the shoe lace before putting it back on. Finally, the overall concept of subtle interaction has an important design consideration with relation to VR and AR; it will undoubtedly sacrifice immersion at the expense of subtleness. All these issues need to be addressed to welcome an era of VR and AR being anytime, anywhere.

References

1 Outram, B. I., Pai, Y. S., Fan, K., Minamizawa, K., & Kunze, K. (2016, October). Any-Orbit: Fluid 6DOF Spatial Navigation of Virtual Environments using Orbital Motion. In Proceedings of the 2016 Symposium on Spatial User Interaction (p. 199). ACM.

2 Pai, Y. S. (2016, October). Physiological Signal-Driven Virtual Reality in Social Spaces. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (pp. 25–28). ACM.

3 Pai, Y. S., Tag, B., Outram, B., Vontin, N., Sugiura, K., & Kunze, K. (2016, July). GazeSim: simulating foveated rendering using depth in eye gaze for VR. In ACM SIGGRAPH 2016 Posters (p. 75). ACM.

4 Pai, Y. S., Outram, B., Vontin, N., & Kunze, K. (2016, October). Transparent Reality: Using Eye Gaze Focus Depth as Interaction Modality. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (pp. 171–172). ACM.

3.11 J!NS MEME – Unobtrusive Smart Eyewear

Yuji Uema (J!NS – Tokyo, JP, yuji-uema@jin-co.com)

JINS CO., LTD released a smart eyewear, JINS MEME in November 2015. It is our new challenge to the field of wearable computing and VR/AR. We initially targeted all consumer but encountered many difficulties regarding the design and benefit of it.

The first issue is social acceptance. Since JINS is an eyewear company, we know that design of eyeglasses matters the most. Therefore, we put our maximum effort to keep JINS MEME hardware looking like a piece of regular eyeglasses. As a result, we designed JINS MEME with bigger temple tips for the battery and electrical circuit. However, we realized that this relatively smaller physical difference compared to regular eyeglasses makes a vast difference in social acceptance. In general, people have accepted and worn eyeglasses more than 700 years. Following this sophisticated design (trend) can be a key factor in social acceptance.

Another challenge is to let people know that we need some practice to get accustomed to these technologies. When it comes to wearable devices and AR/VR devices, people typically do not expect a demand for training. This mindset often makes a gap between user's expectation and benefit of using these technologies. As a result, it sometimes create a negative impression on them.

3.12 Human-centered design of augmentations of social and calm interactions

Kaisa Väänänen (Tampere University of Technology, FI, kaisa.vaananen@tut.fi)

In this research community we can develop advanced technologies that will augment our senses and activities in various life contexts. We should design these augmentations keeping in mind the consequences that these augmentations may have to individuals' life experiences, to their social relationships and even to the humankind (see Figure 1).

An important perspective is that of using augmentation technologies to support social relationships [1]. Such augmentations can enhance sociability remotely or locally. Technologies need to be built for different modalities that enhance both bodily and mental connectedness of people. Studies of social acceptability of the technologies in the real contexts of use are also needed.

Another use of augmenting technologies is to help people to calm down in their hectic everyday lives. One way of supporting this is the actual opposite of adding information to the real world: Diminished reality can remove clutter from the user's surroundings and help people focus on the essentials and also to calm down.

Human-centered design with appropriate user experience goals [2] is a fruitful starting point for such technologically novel augmentations.

Figure 1 Human-centered design ignite talk by Kaisa Vaananen.

References

1. Jarusriboonchai, P., Olsson, T., Malapaschas, A., Väänänen, K. Increase Collocated People's Awareness of Mobile User's Activities: Field Trial of Social Display. Proc. CSCW'16, ACM.
2. Olsson, T., Väänänen-Vainio- Mattila, K., Saari, T., Lucero, A., Arrasvuori, J. Reflections on Experience-Driven Design: a Case Study on Designing for Playful Experiences. Proc. Designing Pleasurable Products and Interfaces, DPPI'13, ACM.

3.13 Socially Acceptable Smart Cameras and AR Glasses

Katrin Wolf (Hamburg University of Applied Sciences – Hamburg, DE, Katrin.Wolf@haw-hamburg.de)

Wearable cameras are nowadays used for adding digital information to our environment through augmented reality (AR) as well as for life logging [1]. We know that lifelog images look best when the camera is worn on the user's head, for example through embedded into glasses [2]. Due to possibilities of automated face and location recognition [3], wearable cameras promise to provide digital information about people we see, meet and talk to, which will be possible right at the moment our wearable camera is capturing them. The lack of information about what smart glasses show about us during a conversation or if the device is recording us while talking makes many people feel uncomfortable when users of wearable cameras and smart glasses are around. With respect to the bystanders' perspective on being captured and computational analyzed, my current research is dedicated to better understand how wearable cameras and their UI should be designed to ensure privacy, to not harm the bystanders' right to be aware what users see about us, to provide ways to object being captured, and in general to foster social acceptability of wearable cameras and AR glasses.

References

1. Katrin Wolf, Albrecht Schmidt, Agon Bexheti and Marc Langheinrich: Lifelogging: You're Wearing a Camera? IEEE Pervasive Computing Magazine, July-Sept. 2014
2. Katrin Wolf, Yomna Abdelrahmen, David Schmid, Tilman Dingler and Albrecht Schmidt: Effects of Camera Position and Media Type on Lifelogging Images. In Proceedings of The 14th International Conference on Mobile and Ubiquitous Multimedia (MUM) 2015.
3. Katrin Wolf, Yomna Abdelrahman, Mathias Landwehr, Geoff Ward and Albrecht Schmidt: How to Browse through my Large Video Data? Face Recognition and Prioritizing for Lifelog Video. In Proceedings of The 15th International Conference on Mobile and Ubiquitous Multimedia (MUM) 2016

4 Workshops

We organized two workshops on hot enabling technologies in VR and AR, to give participants hands-on experiences for later ideation sessions.

4.1 Electric Muscle Stimulation Workshop

Organizer: Pedro Lopes (Hasso Plattner Institute, Potsdam, pedro.lopes@hpi.de)

Current virtual reality technologies focus on vision and sound. However, for better immersion, haptic feedback is needed. Electric Muscle Stimulation (EMS) is an interesting novel mechanism to provide haptics beyond traditional VR and AR applications[1]. Currently, it is still difficult to prototype EMS based systems, as there is a lack of hardware, software and expertise. Pedro Lopes from the Hasso-Plattner-Institut introduced us to his research and open-source effort to make EMS more accessible for research and development [Lopes 2016]. In this hands-on workshop, we first got an introduction into the basics of EMS usage and tried application ideas with a simple open source setup (http://plopes.org/ems/).

References

1. Lopes, Pedro and Pfeiffer, Max and Rohs, Michael and Baudisch, Patrick: Hands-on Introduction to Interactive Electrical Muscle Stimulation. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, 2016.

4.2 SuPerception Workshop

Shunichi Kasahara from Sony's Computer Science Laboratory introduced the term SuPerception which unites super and perception just as he aims at uniting real and artificial perception in his work. He presented examples that include reproducing perceptions and connecting human perceptions through head-worn fisheye cameras and head-mounted displays. Combining the two concepts he enables to be immersed into someone else. Kasahara also showcased his recent work that creates the perception of temporal deformation of the own body in virtual reality by introducing tracking delays or generating prediction of the user's movement[1]. In a shared experience, groups used the parallel eyes system that enables to see three other's perspective video as well as the own perspective through head mounted displays[2].

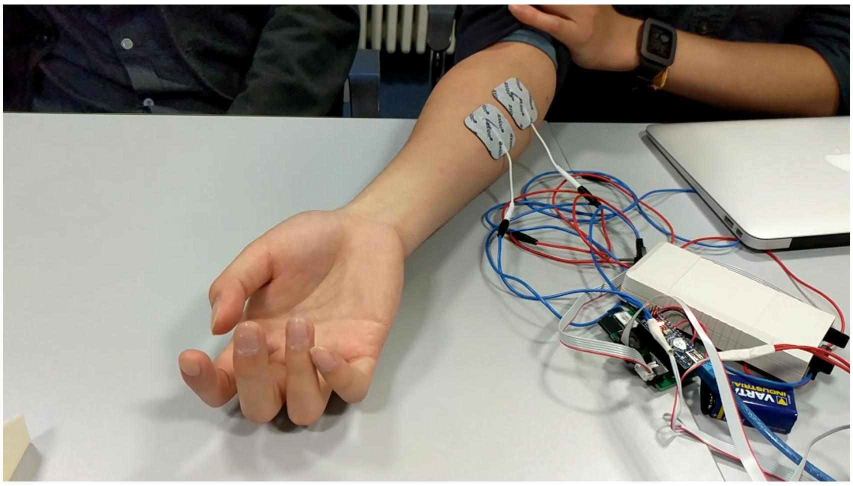


Figure 2 Participants experiencing EMS.

In the future we will see systems that not only digitally alter our perception of reality but provides entirely new abilities that are tightly integrated into our perceptual and motor system. We will be able to zoom into a scene with just a thought or the blink of an eye, fading out parts of our physical environment to focus on a task or instead of learning languages just know them.

References

1 S. Kasahara et al.: MalleableEmbodiment: Changing Sense of Embodiment by Spatial-Temporal Deformation of Virtual Human Body, in Proc. 2017 CHI Conf. Human Factors in Computing Systems (CHI), 2017.

2 S. Kasahara et al.: ParallelEyes: Exploring Human Capability and Behaviors with Paralleled First Person View Sharing, Proc. 2016 CHI Conference on Human Factors in Computing Systems (CHI), 2016, pp. 1561–1572.

5 Thematic Sessions

5.1 Human-Computer Symbiosis

Facilitator: Jonna Häkkilä (University of Lapland, Finland, jonna.hakkila@ulapland.fi)

This thematic session focused on the symbiosis of human and computer. The session started with inspiration talks, and continued with an interactive panel session.

5.1.1 Inspiration Talks

The inspiration talks of the session were given by Susanne Boll, Kaisa Väänänen, and Hans Gellersen, who all gave a short talk of an imaginary superpower of their choice. This was then discussed in respect to the possibilities of human-computer symbiosis. Reflecting on human-computer symbiosis, Susanne Boll asked for the ability to transfer skills to enable people to carry out complex actions on the spot. As an example, she described the scenario of a medical emergency in a remote place, where one was enabled to become a superhero with rescue skills. Through skill transfer, everyone could be enabled to perform a complex

Figure 3 Parallel Eyes session.

medical treatment. As the technical foundation she called for better networked sensors, better reasoning enabled through AI as well as full-body AR, and VR complemented through full-body actuation. Kaisa Väänänen posed the question if technology should focus more on connecting humans through technology instead of human-computer symbiosis. She also challenged the audience by posing self-symbiosis as a challenge. She demanded a superhuman ability that enabled to empty one's own mind, as well as the ability to read the other's mind and body language. Hans Gellersen focused on empathy as a super power. The ability to transfer your point-of-view to somebody else would improve the communication between people and prevent conflicts.

5.1.2 Panel

After discussing human-computer symbiosis with the audience, a facilitated theme panel was organized. Two groups were asked to come forward as panelist teams: Team A (Niels Henze, Susanne Boll, Hans Gellersen) and Team B (Enrico Rukzio, Kaisa Väänänen, Anind Dey). The facilitator of the panel (Jonna Häkkilä) asked the group to discuss controversial questions, and Team A had to always oppose the idea, whereas Team B had to argue for the idea. Both teams had a few minutes to come up with their arguments, and in the meanwhile, the people in the audience had a chance to discuss about the topic with the people next them. The questions or statements given to the panelist teams were as follows: a. In the future, are we able to marry robots? b. In the future, should we be able to replace politicians with AIs? c. In the future, should we have extra robotic arms implanted on us? d. In the future, also my ancestors (like, great-grandfather) will live in my house as (embodied?) avatars. The two teams arguing in favor or opposing the idea resulted very lively discussions touching e.g. social and ethical aspects of human-computer symbiosis.

Figure 4 Panel discussion between Team A and Team B.

5.2 Human Augmentation

Facilitator: Niels Henze (University of Stuttgart, Germany, niels.henze@vis.uni-stuttgart.de)

5.2.1 Inspiration Talks

The inpiration talks in this session were given by Katrin wolf, Mashiko Inami and Thad Starner (see Figure 5), and were followed by a panel discussion with the presenters. Katrin Wolf shared her perspective on sensory augmentation and showed examples of her work on sensory illusion. During the panel discussion, she highlighted, e.g., that humans should stay in control of the level of augmentation. Devices that enable superhuman hearing abilities, for example, must enable users to decide which level of the ability was wanted, from superhearing abilities to blocking the surrounding soundscape. Regarding esthetical questions on human augmentation, she pointed out that technology augmentation that was perceived as uncanny today might be the fashion in the future. For instance, while the third ear the performance artist Stelarc surgically attached to his arm is still considered repelling by many, it might be acceptable in the future. Masahiko Inami showed work from his lab and his driving vision that aims to go from prosthesis to augmentation. Inami draw the link between human evolution which is surpassed by technical evolution. Showing work that equipped users with additional limbs he asked how to control the extra abilities. Thad Starner from Georgia Tech and Google addressed the temporal dimension of human augmentation. He charted the space from passive haptic learning with a delay of hours, to the direct control of the human body by machines with delays less than a millisecond. The main limitation for the augmentation was considered to be the human brain and nervous system itself.

5.2.2 Discussion

The discussion with the audience circled around major challenges that a symbiosis of humans with computers impose. It was asked if augmentation should always add to perception or if it should also reduce experiences. The ability to transmit and share senses on a large scale was discussed, and it was hihglighted that this could ultimately lead to a hive mind

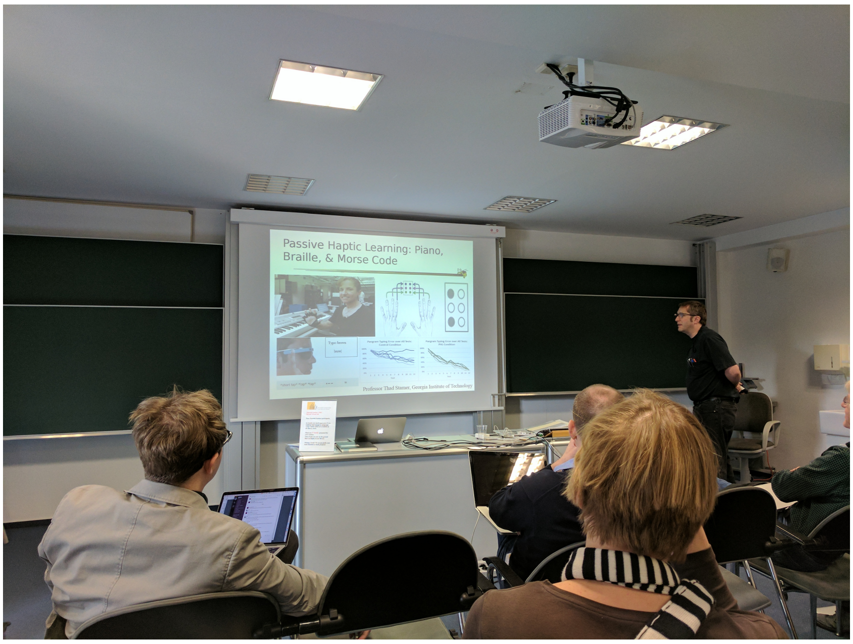


Figure 5 Passive haptic learning by Thad Starner.

society, much like what is depicted by StarTrek's Borg. Social acceptability was seen as a crucial element to lead to or prevent the adoption of the technology. This led to a discussion about fashion and the question if devices that enable superhuman abilities will raise social inequalities to new levels.

5.3 Enabling Technologies

Facilitator: Anind Dey (Carnegie-Mellon University, USA, anind@cs.cmu.edu)

5.3.1 Inspiration Talks

The inpiration talks in the beginning were given by Enrico Rukzio, Florian Michahelles, and Gudrun Klinker. The session on Enabling Technologies focused on the technical issues, means of giving users new skills and experiences, as well as discussing novel or underused methods. Enrico Rukzio started off by discussing on eyewear and eye-based interactions to determine user states. He then continued with scent-based interfaces, which are so far underexplored. Florian Michahelles gave an overview about the industry view on enabling technologies stressing telepresence systems, their progress over the years, and their integration in the future company infrastructure (see Figure 6). He also highlighted the still open challenges related to them. Gudrun Klinker focused on the advances in augmented reality, especially markerless tracking technologies and AR4AR, an automatic calibration system for AR applications. The following discussion focused on the usefulness and applicability of scent-based virtual/augmented environments heading towards more general technologies about extending the human experience away from vision and audio.

Figure 6 View on current industries by Florian Michahelles.

Figure 7 Bodystorming session.

5.3.2 Bodystorming

Followed by the inspiration talks and discussion, an interactive session utilizing a bodystorming co-design method was conducted (see Figure 7). The session involved participants, working in groups, to come up with a future scenario, where technology enabled 'superhuman' power was used in a social setting. These scenarios were then acted by the groups whilst others in the audience. The presented scenarios included, e.g., communicating with thoughts whilst in a business meeting, and semi-automatic behaviour adaptation into the social context.

6 Outcomes

Outcomes of the seminar include plans of joint research projects and fresh perspectives on the attendees' research agenda. In particular, the seminar concluded with a set of challenges for future work:

- Augmented and Virtual Reality research must move from a technology-centric perspective that focuses on computational limitations to a human-centric perspective that considers humans as the most scarce resource.
- With technologies enabling new ways to transfer skills, systems must be tested with real people in real life to identify fundamental challenges and how such systems could transform societies.
- Augmented sports and superhuman sports are an emerging playground for developing and testing new approaches and technologies. As VR and AR technologies become a part of everyday life, work on ethical implications and social acceptance becomes essential.
- Future work must consider a holistic perspective on the user incorporating body and mind. Ultimately, we need methods for describing, visualizing, and interpreting human movement.

With maturing technologies, the community must shift the focus from a very technical approach to a more holistic perspective. Instead of asking how we can build VR and AR systems, we must ask: What do we build? Which new experiences can we create? What are the effects on actual users? How do we cope with users' limited cognitive resources? What will be the implications on the societal level? Participants already started to address these questions. We are looking forward to exciting work that currently emerges from the seminar.

Participants

- Susanne Boll
Universität Oldenburg, DE
- Cedric Carêmel
Takram Design Engineering – Tokyo, JP
- Ashley Colley
University of Lapland – Rovaniemi, FI
- Anind K. Dey
Carnegie Mellon University – Pittsburgh, US
- N. Chloe Eghtebas
Universität Stuttgart, DE
- Hans-Werner Gellersen
Lancaster University, GB
- Nitesh Goyal
Cornell University – Ithaca, US
- Scott Greenwald
MIT – Cambridge, US
- Jonna Häkkilä
University of Lapland – Rovaniemi, FI
- Niels Henze
Universität Stuttgart, DE
- Masahiko Inami
University of Tokyo, JP
- Shunichi Kasahara
Sony CSL – Tokyo, JP
- Gudrun Klinker
TU München, DE
- Kai Kunze
Keio University – Yokohama, JP
- Pedro Lopes
Hasso-Plattner-Institut – Potsdam, DE
- Stephan Lukosch
TU Delft, NL
- Joachim Meyer
Tel Aviv University, IL
- Florian Michahelles
Siemens Corporation – Berkeley, US
- Yun Suen Pai
Keio University – Yokohama, JP
- Thies Pfeiffer
Universität Bielefeld, DE
- Yannick Prié
University of Nantes, FR
- Enrico Rukzio
Universität Ulm, DE
- Thad Starner
Georgia Institute of Technology – Atlanta, US
- Yuji Uema
J!NS – Tokyo, JP
- Kaisa Väänänen
Tampere University of Technology, FI
- Dan Witzner Hansen
IT University of Copenhagen, DK
- Katrin Wolf
HAW – Hamburg, DE
- Eva Wolfangel
Stuttgart, DE

Report from Dagstuhl Seminar 17071

Computer-Assisted Engineering for Robotics and Autonomous Systems

Edited by

Erika Abraham[1], Hadas Kress-Gazit[2], Lorenzo Natale[3], and Armando Tacchella[4]

1 **RWTH Aachen, DE,** abraham@cs.rwth-aachen.de
2 **Cornell University – Ithaca, US,** hadaskg@cornell.edu
3 **Italian Institute of Technology – Genova, IT,** lorenzo.natale@iit.it
4 **University of Genova, IT,** armando.tacchella@unige.it

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 17071 "Computer-Assisted Engineering for Robotics and Autonomous Systems". This seminar brought together researchers from three distinct communities – Robotics, Model-driven Software Engineering, and Formal Methods – to discuss the path towards creating safe and verifiable autonomous systems.

Seminar February 12–17, 2017 – http://www.dagstuhl.de/17071
1998 ACM Subject Classification D.2.4 Software/Program Verification, I.2.9 Robotics
Keywords and phrases analysis, artificial intelligence, autonomous systems, computer-aided software development, domain-specific languages, formal methods, model-driven software engineering, robotics, safety, synthesis, testing, verification
Digital Object Identifier 10.4230/DagRep.7.2.48
Edited in cooperation with Rüdiger Ehlers

1 Executive Summary

Hadas Kress-Gazit
Erika Abraham
Lorenzo Natale
Armando Tacchella

This seminar focused on *autonomous systems*, and more specifically robots, that operate without, or with little, external supervision. For these systems to be integrated into society, it is highly important to make sure that they are functionally safe. *Formal Methods* are techniques adopted in engineering for the verification of software and hardware systems. As models are a basic requirement for the formal analysis of systems, *Model-driven Software Engineering* plays an important role to enable the application of *Formal Methods*. Though autonomous systems are increasingly involved in our everyday life, both exact formalizations of safe functionality (standards, what we want to be confident in) and methods to achieve confidence (methodologies, how we get confident in the properties we want to assure) are still scarce.

This seminar brought together experts in *Artificial Intelligence* and *Robotics*, *Model-driven Software Engineering*, and *Formal Methods*. It included researchers from academia as well

Computer-Assisted Engineering for Robotics and Autonomous Systems, *Dagstuhl Reports*, Vol. 7, Issue 2, pp. 48–63
Editors: Erika Abraham, Hadas Kress-Gazit, Lorenzo Natale, and Armando Tacchella
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

as from industry. The following list summarizes high-level themes that emerged from the seminar:

- Dealing with highly complex systems, it is difficult to verify or even model all aspects of the system, therefore focusing effort on efficient falsification rather than costly verification can be highly impactful for industrial applications.
- The community can and should leverage results and systems built for different robotic competitions to reason about possible requirements and techniques to verify/falsify them. These competitions include the DARPA robotics challenge, the Amazon picking challenge, different leagues in Robocup, etc. Creating benchmarks based on these competitions will enable progress in verification of autonomous systems.
- Creating small interdisciplinary teams that include people from formal methods, robotics and model based design that tackle small yet realistic problems, possibly inspired by industrial applications, will help formalize the language of requirements, models and verification techniques that will have an impact on autonomous systems.

2 Table of Contents

3 Overview of Talks

3.1 The Power of Satisfiability Checking

Erika Abraham (RWTH Aachen, DE)

Joint work of Erika Abraham and Gereon Kremer
Main reference Erika Ábrahám, Gereon Kremer, "Satisfiability Checking: Theory and Applications", in Proc. of the 14th Int'l Conf. on Software Engineering and Formal Methods (SEFM 2016), LNCS, Vol. 9763, pp. 9–23, Springer International Publishing, 2016.
URL http://dx.doi.org/10.1007/978-3-319-41591-8_2

Satisfiability checking aims to develop algorithms and tools for checking the satisfiability of existentially quantified logical formulas. For propositional logic, in the late '90s impressive progress was made towards practically applicable solutions, resulting in powerful SAT solvers. Driven by this success, a new line of research started to enrich propositional SAT solving with solver modules for different theories. Nowadays, sophisticated SAT-modulo-theories (SMT) solvers are available for, e.g., equality logic with uninterpreted functions, bit-vector arithmetic, array theory, floating point arithmetic, and real and integer arithmetic. SAT and SMT solvers are now at the heart of many techniques for the analysis of programs and probabilistic, timed, hybrid and cyber-physical systems, for test-case generation, for solving large combinatorial problems and complex scheduling tasks, for product design optimisation, planning and controller synthesis, just to mention a few well-known areas.

In this talk we gave a short introduction to the theoretical foundations of satisfiability checking, mentioned some of the most popular tools, and discussed the successful embedding of SMT solvers in different technologies.

3.2 Model-Driven Control Software / System Design for Robotic Systems

Jan Broenink (University of Twente, NL)

In dealing with system architectures for robotic and automation systems, it is crucial to consider the total system (machine, control, software and I/O), because the dynamics of the machine influences the robot software. Therefore, we use appropriate Models of Computation and tools, namely bond graphs for the machine part, dataflow diagrams for the algorithm / software parts. Via meta-models, these formalisms are related. This allows for a structured approach for designing the architecture of the robotic system. The design work is done as a stepwise refinement process, whereby each step is verified via simulation, yielding shorter design time, and a better quality product. The tools use templates and pass model-specific information between each other via parameterised tokens in the generated, high-level code, to get a better separation of design steps. This allows for better quality of the models and more reuse, thus enhancing the efficiency of model-driven design for the (industrial) end user. This approach is illustrated with two case studies: the control stack for a mobile robot, manipulating blocks, and on incorporating safety layers in the embedded control system architecture.

3.3 Safety Cases. Arguing the Safety of Autonomous Systems

Simon Burton (Robert Bosch GmbH – Stuttgart, DE)

This talk introduced the topic of safety cases for arguing the safety of autonomous systems. Examples are given for where existing standards do not provide sufficient guidance to demonstrate certain properties of autonomous systems and therefore require a justification from "first principles". The Goal Structuring Notation is described as a means of formulating and communication such argumentation structures. A roadmap for how to extend these concepts in combination with model-based Systems Engineering and formal methods is presented to motivate future research and encourage collaboration between these domains.

3.4 Computer-Assisted Engineering for Robotics and Autonomous Systems: Verification Techniques That (May) Work in Practice

Kerstin I. Eder (University of Bristol, GB)

Joint work of Dejanira Araiza-Illan, David Western, Pjotr Trojanek, Anthony G. Pipe, Arthur Richards, Kerstin I. Eder

This presentation is focused on practical techniques for the verification of autonomous systems. Because no single technique is adequate to cover a whole system in practice, the use of a variety of techniques is proposed, including formal and state-of-the-art simulation-based, to address verification needs in autonomous system design.

At the code level, re-implementing three well-known robot navigation algorithms in SPARK enables formal verification to establish freedom from run-time errors without performance penalties when compared to implementations in C/C++ [1]. This shows that selecting a programming language designed for software-reliability leads to significant advantages when it comes to establishing code correctness.

At the design level, an assertion-based approach is proposed to verify control system designs with respect to high-level requirements, such as stability, combining simulation-based techniques with automatic theorem proving [2]. Requirements are first formalized as properties over the signals in the Simulink model using Simulink blocks that then become part of the Simulink model. The so extended Simulink model is then automatically translated into Why3 theories and proof goals for formal verification using SMT-based theorem provers. A case study that illustrates how stability can be decomposed from a single high-level requirement into a set of sub-requirements to be implemented as assertions in Simulink is discussed [3], together with the advantages of combining assertion-checks performed during simulation with automatic theorem proving performed at system design time.

Coverage-Driven Verification (CDV) is as a systematic, goal directed simulation-based verification method that is capable of exploring systems of realistic detail under a broad range of environment conditions, providing a high degree of automation. I will illustrate the benefits of CDV, functional and situation coverage [4] together with model-based [5] as well as intelligent, agent-based test generation techniques [6] on the example of code used in robots that directly interact with humans.

I conclude my presentation with a brief discussion of the challenges in this area: specification, automation, combination of techniques and using AI for verification and validation.

Acknowledgement. The research presented is based on collaborations within the EPSRC funded projects "Robust Integrated Verification of Autonomous Systems" (EP/J01205X/1) and "Trustworthy Robotic Assistants" (EP/K006320/1).

References

1 Piotr Trojanek and Kerstin Eder. Verification and Testing of Mobile Robot Navigation Algorithms: A Case Study in SPARK. In *Proc. International Conference on Intelligent Robots and Systems (IROS)*, 2014.

2 Dejanira Araiza-Illan, Kerstin Eder, and Arthur Richards. Formal verification of control systems' properties with theorem proving. In *2014 UKACC International Conference on Control (CONTROL)*, Loughborough, UK, July 2014.

3 Dejanira Araiza-Illan, Kerstin Eder, and Arthur Richards. Verification of Control Systems Implemented in Simulink with Assertion Checks and Theorem Proving: A Case Study. In *European Control Conference (ECC)*, Linz, Austria, 2015.

4 Dejanira Araiza-Illan, David Western, Anthony Pipe, and Kerstin Eder. Coverage-Driven Verification: An Approach to Verify Code for Robots that Directly Interact with Humans. In *Haifa Verification Conference*. Haifa, Israel, 2015. DOI: 10.1007/978-3-319-26287-1_5.

5 Dejanira Araiza-Illan, David Western, Anthony G. Pipe, and Kerstin Eder. Systematic and Realistic Testing in Simulation of Control Code for Robots in Collaborative Human-Robot Interactions. In *Towards Autonomous Robotic Systems (TAROS)*. June 2016.

6 Dejanira Araiza-Illan, Anthony G. Pipe, and Kerstin Eder. Intelligent agent-based stimulation for testing robotic software in human-robot interactions. In *Proceedings of the 3rd Workshop on Model-Driven Robot Software Engineering*, MORSE'16, pages 9–16, New York, NY, USA, 2016. ACM.

3.5 Towards Best-Effort Autonomy

Rüdiger Ehlers (Universität Bremen, DE)

Joint work of Rüdiger Ehlers, Salar Moarref, Ufuk Topcu
Main reference Rüdiger Ehlers, Salar Moarref, Ufuk Topcu, "Risk-averse control of Markov decision processes with ω-regular objectives", in Proc. of the 55th IEEE Conference on Decision and Control (CDC 2016), pp. 426–433, IEEE, 2016.
URL http://dx.doi.org/10.1109/CDC.2016.7798306
URL http://progirep.github.io/ramps/

Highly autonomous systems degrade in performance over time, need to work correctly in off-nominal conditions, and need to adapt without the help of a human operator. We do not always know in advance of the system's deployment how they are degrading in the long run, and not all possible degradation scenarios can be covered in a systematic system engineering process. To counter this problem, we could synthesize adapted control strategies at runtime, using action failure probabilities inferred from observed data. However, classical policy synthesis techniques for ω-regular specifications yield no policy in case of inevitable eventual violation of the specification. We present an approach to mitigate this problem for omega-regular specifications and environments that can be modelled as Markov decision processes.

3.6 Provably Safe Collision Avoidance in Dynamic Environments

Christian Heinzemann (Robert Bosch GmbH – Stuttgart, DE)

For many applications of autonomous robots in intralogistics and mobile service robotics, it is an absolute must to guarantee that the robot will not cause harm to its environment. This particularly includes that the robot must not cause collisions with moving obstacles such as humans or animals. Guaranteeing collion-free motion of autonomous systems is increasingly hardened by the fact that these systems increasingly operate in shared, open-context environments. In such environments, the robot operates in the same space as the humans and we as the developers do not know all contexts in which the system will have to operate during its runtime. In particular, we will often not know how the environment looks like and which kinds of obstacles the system will face. To this end, an approach for guaranteeing provably safe motion of mobile robots is necessary. The main safety concept being adopted therefore is passive safety [1], requiring that the robot is not moving when a collision with an obstacle happens. The existing approaches either make the optimistic assumption of knowing the future behavior of any obstable [2, 3], which is unrealistic for humans, or they make rather conservative assumptions about obstacles [4, 5, 6, 7] that significantly decrease the robot's performance. The latter is true particularly in cases where many obstacles are in the robot's environment and where these obstacles are relatively near to the robot, for example, when moving through an area populated by humans in a city center, airport, or train station. Probabilistic approaches to collision avoidance [8, 9] improve the performance but cannot give the necessary safety guarantees that we need for heavy robots used, for example, in intralogistics.

In this talk, I give briefly characterize the problem of collision avoidance to be solved for mobile robots and discuss in more detail why the problem is not solved sufficiently by existing approaches. A possible trail for future works could be online verification approaches based on reachability analysis [10] that use models to overapproximate the space that an obstacle will occupy at the end of a planning period of the reactive obstacle avoidance algorithms. I conclude by summarizing the key challenges that need to be solved for the approaches.

References

1. K. Macek, D. Vasquez, T. Fraichard, R. Siegwart: Towards Safe Vehicle Navigation in Dynamic Urban Scenarios. In: Automatica 50, Issue 3–4, pp. 184–194, 2009.
2. L. Martinez-Gomez and T. Fraichard: Collision avoidance in dynamic environments: An ICS-based solution and its comparative evaluation. In: Proceedings of ICRA 2009.
3. F. Large, C. Laugier, and Z. Shiller: Navigation Among Moving Obstacles Using the NLVO: Principles and Applications to Intelligent Vehicles. In: Autonomous Robots 19(2), 2005.
4. S. Bouraine, T. Fraichard, O. Azouaoui, H. Salhi: Passively safe partial motion planning for mobile robots with limited field-of-views in unknown dynamic environments. In: Proceedings of ICRA 2014, pp. 3576–3582, 2014.
5. S. Mitsch, K. Ghorbal, and A. Platzer: On Provably Safe Obstacle Avoidance for Autonomous Robotic Ground Vehicles. Proceedings of Robotics: Science and Systems, 2013.
6. M. Zhang, X. Zhang: Formally verifying navigation safety for ground robots. 2016 IEEE International Conference on Mechatronics and Automation.
7. C. Dabadie, S. Kaynama, C.J. Tomlin: A practical reachability-based collision avoidance algorithm for sampled-data systems: Application to ground robots. Proceedings of IROS 2014, pp. 4161–4168, 2014.

8 A. Bautin, L. Martinez-Gomez, T. Fraichard: Inevitable Collision States: A probabilistic perspective. In: Proceedings of ICRA 2010.

9 D. Althoff, M. Althoff, D. Wollherr, M. Buss: Probabilistic collision state checker for crowded environments. In: Proceedings of ICRA 2010.

10 M. Althoff, J.M. Dolan: Online Verification of Automated Road Vehicles Using Reachability Analysis. IEEE Transactions on Robotics (30), pp. 903–918, 2014.

3.7 Heteronomous Systems They are, Let's Face it.

Holger Hermanns (Universität des Saarlandes, DE)

Joint work of Raimund Dachselt, Holger Hermanns

Heteronomy refers to actions that are influenced by forces outside the individual. Autonomy is the opposite. For good reason, cars were originally called automobiles (and in some languages they still are). They give autonomy to people. So, what is an autonomous automobile?

In this talk I will argue that the currently acclaimed vision of fully autonomous systems is nothing but a trend towards heteronomy. This puts computer-assistance for heteronomous system design into a different perspective. I will elaborate on this perspective, and will discuss research challenges directly resulting from this.

3.8 GenoM3 Templates: from Middleware Independence to Formal Models Synthesis

Felix Ingrand (LAAS – Toulouse, FR)

Joint work of Mohammed Foughali, Félix Ingrand, Anthony Mallet
Main reference Mohammed Foughali, Félix Ingrand, Anthony Mallet, "GenoM3 Templates: from Middleware Independence to Formal Models Synthesis", Rapport LAAS no. 17022. 2017.
URL https://hal.laas.fr/hal-01457881

GenoM is an approach to develop robotic software components, which can be controlled, and assembled to build complex applications. Its latest version, GenoM3, provides a template mechanism which is versatile enough to deploy components for different middleware without any change in the specification and user code. But this same template mechanism also enables us to automatically synthesize formal models (for two Validation and Verification frameworks) of the final components. We present and illustrate our approach on a real deployed example of a drone flight controller for which we prove offline real-time properties, and an outdoor robot for which we synthesize a controller to perform runtime verification.

This work was supported in part by the EU CPSE Labs project funded by the H2020 program under grant agreement No. 644400.

3.9 Synthesis of Shared Control Protocols with Provable Safety and Performance Guarantees

Nils Jansen (Univ. of Texas at Austin, US)

Joint work of Nils Jansen, Murat Cubuktepe, Ufuk Topcu
Main reference Nils Jansen, Murat Cubuktepe, Ufuk Topcu, "Synthesis of Shared Control Protocols with Provable Safety and Performance Guarantees", in Proc. of the 2017 American Control Conference (ACC'17), preprint available at arXiv:1610.08500v1 [cs.RO], 2017.
URL https://arxiv.org/abs/1610.08500v1

We formalize synthesis of shared control protocols with correctness guarantees for temporal logic specifications. More specifically, we introduce a modeling formalism in which both a human and an autonomy protocol can issue commands to a robot towards performing a certain task. These commands are blended into a joint input to the robot. The autonomy protocol is synthesized using an abstraction of possible human commands accounting for randomness in decisions caused by factors such as fatigue or incomprehensibility of the problem at hand. The synthesis is designed to ensure that the resulting robot behavior satisfies given safety and performance specifications, e.g., in temporal logic. Our solution is based on nonlinear programming and we address the inherent scalability issue by presenting alternative methods. We assess the feasibility and the scalability of the approach by an experimental evaluation.

3.10 A storm is Coming: A Modern Probabilistic Model Checker

Joost-Pieter Katoen (RWTH Aachen, DE)

Joint work of Christian Dehnert, Sebastian Junges, Joost-Pieter Katoen, Matthias Volk
Main reference Christian Dehnert, Sebastian Junges, Joost-Pieter Katoen, Matthias Volk, "A storm is Coming: A Modern Probabilistic Model Checker", in Proc. of the 29th Int'l Conf. on Computer Aided Verification (CAV'17), LNCS, Vol. 10427, pp. 592-600, Springer, 2017.
URL http://dx.doi.org/10.1007/978-3-319-63390-9_31
URL https://moves-rwth.github.io/storm/

In the last five years, we have developed our in-house probabilistic model checker with the aim to have an easy-to-use platform for experimenting with new verification algorithms, richer probabilistic models, algorithmic improvements, different modeling formalism, various new features, and so forth. Although open-source probabilistic model checkers do exist, most are not flexible and modular enough to easily support this. Our efforts have led to a toolkit with mature building bricks with simple interfaces for possible extensions, and a modular set-up. It comprises about 100,000 lines of `C++` code. The time has come to make this toolkit available to a wider audience: this paper presents storm.

Like its main competitors PRISM, MRMC, and iscasMC, storm relies on numerical and symbolic computations. It does not support discrete-event simulation, known as statistical model checking. The main characteristic features of storm are:

- it supports *various native input formats*: the PRISM input format, generalized stochastic Petri nets, dynamic fault trees, and conditioned probabilistic programs. This is not just providing another parser; state-space reduction and generation techniques as well as analysis algorithms are partly tailored to these modeling formalisms;

- in addition to Markov chains and MDPs, it supports *Markov automata*, a model containing probabilistic branching, non-determinism, and exponentially distributed delays;
- it can do *explicit state* and *fully symbolic* (BDD-based) model checking as well as a *mixture* of these modes;
- it has a *modular* set-up, enabling the easy exchange of different solvers and distinct decision diagram packages; its current release supports about 15 solvers, and the BDD packages `CUDD` [1] and multi-threaded `Sylvan` [2];
- it provides a *Python API* facilitating easy and rapid prototyping of other tools using the engines and algorithms in storm;
- it provides the following functionalities under one roof: the synthesis of counterexamples and permissive schedulers (both MILP- and SMT-based), game-based abstraction of infinite-state MDPs, efficient algorithms for conditional probabilities and rewards, and long-run averages on MDPs;
- its performance in terms of verification speed and memory footprint on the PRISM benchmark suite is mostly better compared to PRISM

Although many functionalities of PRISM are covered by storm, there are significant differences. storm does not support LTL model checking and does not support the PRISM features: probabilistic timed automata, multi-objective model checking, and an equivalent of PRISM's "hybrid" engine (a crossover between full MTBDD and storm's "hybrid" engine), a fully symbolic engine for continuous-time models, statistical model checking, and the analysis of stochastic games as in PRISM-GAMES.

References

1. Fabio Somenzi: CUDD: CU Decision Diagram package release 3.0.0, http://vlsi.colorado.edu/~fabio/CUDD/ (2016)
2. Tom van Dijk and Jaco van de Pol: Sylvan: Multi-Core Decision Diagrams. *Tools and Algorithms for the Construction and Analysis of Systems – 21st International Conference* (TACAS 2015). pp. 677–691, 2014.

3.11 High-Level Verifiable Robotics

Hadas Kress-Gazit (Cornell University – Ithaca, US)

Joint work of Gangyuan Jing, Tarik Tosun, Mark Yim, Scott Hamill, Jon DeCastro, Kai Weng Wong, Hadas Kress-Gazit
URL http://verifiablerobotics.com/

In this talk I gave a quick overview of different projects in my lab in which we have used LTL synthesis and verification techniques to automatically create provably-correct robot controllers. I finished the talk with a provocative question on what is the role of formal verification and synthesis in the era of learning-based robotics.

3.12 (Learning to) Learn to Control

Jan Kretinsky (TU München, DE)

Joint work of P. Ashok, T. Brazdil, K. Chatterjee, M. Chmelik, P. Daca, A. Fellner, V. Forejt, T. Henzinger, J. Kretinsky, M. Kwiatkowska, T. Meggendorfer, D. Parker, T. Petrov, V. Toman, M. Ujma
Main reference Tomás Brázdil, Krishnendu Chatterjee, Martin Chmelik, Vojtech Forejt, Jan Křetínský, Marta Z. Kwiatkowska, David Parker, Mateusz Ujma, "Verification of Markov Decision Processes Using Learning Algorithms", in Proc. of the 12th International Symposium on Automated Technology for Verification and Analysis (ATVA 2014), LNCS, Vol. 8837, pp. 98–114, Springer, 2014.
URL http://dx.doi.org/10.1007/978-3-319-11936-6_8

On the one hand, formal verification methods provide hard guarantees on analysis results, but do not scale well and are often hard to use. On the other hand, machine learning comes with weak or no guarantees, but scales well and can provide more understandable solutions. In this talk, we show several examples how these approaches can be combined and the best of the two worlds achieved. We demonstrate this on controller synthesis [1,2] and controller representation [3] in the setting of Markov decision processes and comment on extensions to games [4].

References

1 Tomás Brázdil, Krishnendu Chatterjee, Martin Chmelik, Vojtech Forejt, Jan Křetínský, Marta Z. Kwiatkowska, David Parker, and Mateusz Ujma. *Verification of Markov decision processes using learning algorithms*. In *ATVA*, pages 98–114. Springer, 2014.
2 P. Ashok, K. Chatterjee, P. Daca, J. Křetínský, T. Meggendorfer. *Value Iteration for Long-run Average Reward in Markov Decision Processes*. Submitted.
3 T. Brazdil, K. Chatterjee, M. Chmelik, A. Fellner, J. Křetínský. *Counterexample Explanation by Learning Small Strategies in Markov Decision Processes*. In *CAV (1)*, pages 158–177. Springer, 2015.
4 T. Brazdil, K. Chatterjee, J. Křetínský, V. Toman. *Strategy Representation by Decision Trees in Reactive Synthesis*. Submitted.

3.13 Optimizing the Performance of Robots in Production Logistics Scenarios

Gerhard Lakemeyer (RWTH Aachen, DE)

Joint work of Tim Niemueller, Erika Abraham, Gerhard Lakemeyer
Main reference Frederik Zwilling, Tim Niemueller, Gerhard Lakemeyer, "Simulation for the RoboCup Logistics League with Real-World Environment Agency and Multi-level Abstraction", in RoboCup 2014: RoboCup 2014: Robot World Cup XVIII, LNCS, Vol. 8992, pp. 220–232, Springer, 2015.
URL http://dx.doi.org/10.1007/978-3-319-18615-3_18
URL https://www.fawkesrobotics.org/media/publications/llsf-sim-rc2014.pdf

We consider the problem of optimizing the decision making of mobile robots managing the supply chain in a semi-structured factory setting. To keep things manageable and comprehensible we focus on a game-like environment provided by the Robocup Logistics League (RCLL). While the RCLL has been around for a number of years, there has been little progress in optimizing the performance of the robots. In order to make progress in a more principled way we recently joined forces with Erika Abraham's group with the aim of applying SMT techniques to this problem. In this talk I will mainly focus on describing the problems and challenges the RCLL raises and advertise the simulation-based variant of the RCLL as a possible benchmark to develop and test formal methods in robotics. I will also briefly outline our approach and the first steps we have taken to address the problem using SMT.

3.14 Artificial Intelligence Planning and Robotics and Autonomous Systems

Daniele Magazzeni (King's College London, GB)

Joint work of Daniele Magazzeni, Michael Cashmore, Maria Fox, Derek Long
Main reference Michael Cashmore, Maria Fox, Derek Long, Daniele Magazzeni, "A Compilation of the Full PDDL+ Language into SMT", in Proc. of the 26th Int'l Conf. on Automated Planning and Scheduling (ICAPS 2016), pp. 79–87, AAAI Press, 2016.
URL https://www.aaai.org/ocs/index.php/ICAPS/ICAPS16/paper/view/13101
URL http://kcl-planning.github.io/ROSPlan/

AI Planning is about determining actions before doing them, anticipating the things that will need to be done and preparing for them. Planners use domain-independent heuristics to guide the search in huge state spaces, in order to find a plan that achieves the goal while satisfying numerical and temporal constraints and optimising a given metric. Planning for Robotics and Autonomous Systems requires rich models to capture complex dynamics as well as the uncertain and evolving environment, scalable planning techniques and robust methods of execution. PDDL+ is the formalism used in planning to describe hybrid systems, and allows the modelling of the differential equations governing the continuous behaviour of systems. This talk provides an overview of how PDDL+ can be used to model complex domains; presents a new PDDL+ planner based on SMT and the ROSPlan framework for planning with ROS; highlights some open challenges on the integration between task and motion planning.

3.15 Human-Robot Collaboration – Industrial Applications and Open Challenges

Björn Matthias (ABB AG Forschungszentrum – Ladenburg, DE)

Main reference Björn Matthias, "Risk Assessment for Human-Robot Collaborative Applications", Workshop on Physical Human-Robot Collaboration: Safety, Control, Learning and Applications at the IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS 2015),2015.
URL https://www.researchgate.net/publication/282778774_Risk_Assessment_for_Human-Robot_Collaborative_Applications

This contribution seeks to identify some important gaps in present methodology in the deployment of industrial robots in applications of human-robot collaboration (HRC). The drivers and enablers for deployment of industrial robots and of HRC in industrial practice are summarized. Safety of machinery, as called out for example in the European Machinery Directive, is introduced as a necessary boundary condition to fulfill in applications of industrial robots. A brief overview of the relevant standards to be followed is given. The basic types of collaborative operation of industrial robots are summarized, describing the specific protection schemes for each. The challenges in planning and commissioning collaborative applications in industrial production are considered in more detail. This allows the identification of the present lack of methodology and tools to support the economical and safety-rated deployment of applications using HRC. The resulting research questions address these and other issues associated with the future of industrial robots and their applications.

3.16 A Competition on Formal Methods for Robotics

Vasumathi Raman (Zoox Inc., US)

Joint work of Vasumathi Raman, Scott C. Livingston
Main reference Vasumathi Raman, "The 2016 Formal Methods for Robotics Challenge [Competitions]", IEEE Robotics & Automation Magazine, Vol. 23(3), pp. 24–25, 2016.
URL http://dx.doi.org/10.1109/MRA.2016.2587958
URL https://fmrchallenge.org/

Formal methods refers broadly to techniques for the verification and automatic synthesis of transition systems that satisfy desirable properties exactly or within some statistical tolerance. Though historically developed for concurrent software, recent work has brought these methods to bear on motion planning in robotics. Challenges specific to robotics, such as uncertainty and real-time constraints, have motivated extensions to existing methods, as well as entirely novel treatments. However, when compared with other areas within robotics research, demonstrations of formal methods have been surprisingly small-scale. In this talk, I propose a robotics challenge that seeks to motivate advancement of the state of the art toward practical realization. The challenge is organized into three problem domains: arbitrary dimensional chains of integrators, traffic networks with Dubins cars, and factory cart clearing.

3.17 Specification: the Biggest Bottleneck in Formal Methods and Autonomy

Kristin Yvonne Rozier (Iowa State University, US)

Main reference Kristin Yvonne Rozier, "Specification: The Biggest Bottleneck in Formal Methods and Autonomy", in Proc. of the 8th Working Conf. on Verified Software: Theories, Tools, and Experiments (VSTTE'16), LNCS, Vol. 9971, pp. 1–19, Springer, 2016.
URL http://dx.doi.org/10.1007/978-3-319-48869-1_2

Advancement of autonomous systems stands on the shoulders of formal methods, which make possible the rigorous safety analysis autonomous systems require. An aircraft cannot operate autonomously unless it has design-time reasoning to ensure correct operation of the autopilot and runtime reasoning to ensure system health management, or the ability to detect and respond to off-nominal situations. Formal methods are highly dependent on the specifications over which they reason; there is no escaping the "garbage in, garbage out" reality. Specification is difficult, unglamorous, and arguably the biggest bottleneck facing verification and validation of autonomous systems.

We examine the outlook for formal specification, and highlight the on-going challenges of specification, from design-time to runtime. We exemplify these challenges for specifications in

Linear Temporal Logic (LTL) though the focus is not limited to that specification language. We pose challenge questions for specification that will shape both the future of formal methods, and our ability to more automatically verify and validate autonomous systems of greater variety and scale. We call for further research into LTL Genesis.

3.18 Development and Adoption of Model-Based Tools in Robotics

Christian Schlegel (Hochschule Ulm, DE)

Joint work of Christian Schlegel, Alex Lotz, Matthias Lutz, Dennis Stampfer
Main reference Dennis Stampfer, Alex Lotz, Matthias Lutz, Christian Schlegel, "The SmartMDSD Toolchain: An Integrated MDSD Workflow and Integrated Development Environment (IDE) for Robotics Software", Special Issue on Domain-Specific Languages and Models in Robotics, Journal of Software Engineering for Robotics (JOSER), 7(1), pp. 3–19, ISSN: 2035-3928, Open Journal Systems, 2016.
URL https://joser.unibg.it/index.php?journal=joser&page=article&op=view&path%5B%5D=91

We aim at making the development of better quality robot systems much less an effort by the means of model-driven tooling. This talk is about how to compose complex robotic software systems out of software building blocks and we advocate for moving from just source-code level integration towards model-driven composition with explicated properties. We consider the full stack from low level control over the task sequencing level up to the mission level. The challenge is to adhere to the principles of separation of concerns while at the same time, you need to package different concerns into structures such that these fit the views of e.g. component developers, system integrators and even the robots at run-time themselves. This talk underpins these ideas by the example of the matured model-driven SmartSoft/SmartMDSD approach and tooling. At various levels, there are hooks in the software engineering tools and in the run-time execution system where (formal) methods (e.g. for verification) could assist the different players in their different roles including the robot itself in better doing their jobs.

3.19 How Safe is Your Autonomous Robot? (A Tale of Courage, Passion, and Perspiration)

Armando Tacchella (University of Genova, IT)

In this work we consider the problem of checking safety in autonomous agents at the deliberative level. The interaction between the agent and the environment is modelled as a Markov decision process and it is assumed that control policies are learned using model-free approximate dynamic programming, best known as reinforcement learning (RL). Models and policies inferred during RL are combined to obtain discrete time Markov chains which can then be subject to verification and repair against probabilistic temporal logic properties. In particular, we consider repair both as an off-line strategy and an on-line technique to supplement execution monitoring with policy-mending capabilities. The approach is studied in the context of a standing-up task for a simple but nontrivial humanoid robot.

Participants

- Erika Abraham
RWTH Aachen, DE
- Jan Broenink
University of Twente, NL
- Simon Burton
Robert Bosch GmbH – Stuttgart, DE
- Pablo Bustos
University of Extremadura – Cáceres, ES
- Alessandro Cimatti
Bruno Kessler Foundation – Trento, IT
- Manuel Alcino Cunha
INESC TEC and University of Minho – Braga, PT
- Kerstin I. Eder
University of Bristol, GB
- Rüdiger Ehlers
Universität Bremen, DE
- Samira Farahani
MPI-SWS – Kaiserslautern, DE
- Stefano Ghidoni
University of Padova, IT
- Christoffer R, Heckman
University of Colorado – Boulder, US
- Christian Heinzemann
Robert Bosch GmbH – Stuttgart, DE
- Holger Hermanns
Universität des Saarlandes, DE
- Felix Ingrand
LAAS – Toulouse, FR
- Nils Jansen
Univ. of Texas at Austin, US
- Joost-Pieter Katoen
RWTH Aachen, DE
- Hadas Kress-Gazit
Cornell University – Ithaca, US
- Jan Kretinsky
TU München, DE
- Morteza Lahijanian
University of Oxford, GB
- Gerhard Lakemeyer
RWTH Aachen, DE
- Ratan Lal
Kansas State University – Manhattan, US
- Martin Leucker
Universität Lübeck, DE
- Ingo Lütkebohle
Robert Bosch GmbH – Stuttgart, DE
- Daniele Magazzeni
King's College London, GB
- Björn Matthias
ABB AG Forschungszentrum – Ladenburg, DE
- Lorenzo Natale
Italian Institute of Technology – Genova, IT
- Sandeep Neema
DARPA – Arlington, US
- Petter Nilsson
University of Michigan – Ann Arbor, US
- Shashank Pathak
Technion – Haifa, IL
- Subramanian Ramamoorthy
University of Edinburgh, GB
- Vasumathi Raman
Zoox Inc., US
- Kristin Yvonne Rozier
Iowa State University, US
- Christian Schlegel
Hochschule Ulm, DE
- Maria Svorenova
University of Oxford, GB
- Armando Tacchella
University of Genova, IT
- Sebastian Wrede
Universität Bielefeld, DE

Report from Dagstuhl Seminar 17072

Applications of Topology to the Analysis of 1-Dimensional Objects

Edited by

Benjamin Burton[1], Maarten Löffler[2], Carola Wenk[3], and Erin Moriarty Wolf Chambers[4]

1 **The University of Queensland, AU,** bab@maths.uq.edu.au
2 **Utrecht University, NL,** m.loffler@uu.nl
3 **Tulane University, US,** cwenk@tulane.edu
4 **St. Louis University, US,** echambe5@slu.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 17072 "Applications of Topology to the Analysis of 1-Dimensional Objects".

Seminar February 12–17, 2017 – http://www.dagstuhl.de/17072
1998 ACM Subject Classification I.3.5 Computational Geometry and Object Modeling
Keywords and phrases curves, graph drawing, homotopy, knot theory
Digital Object Identifier 10.4230/DagRep.7.2.64
Edited in cooperation with Hsien-Chih Chang

1 Executive Summary

Benjamin Burton
Maarten Löffler
Carola Wenk
Erin Moriarty Wolf Chambers

Description of the seminar

One-dimensional objects embedded in higher-dimensional spaces are one of the most natural phenomena we encounter: ranging from DNA strands to roads to planetary orbits, they occur at all granularities throughout the sciences. Computer-assisted analysis of one-dimensional data is now standard procedure in many sciences; yet the underlying mathematics are not always well understood, preventing the most powerful analytical tools from being used.

Adding to the confusion, one-dimensional objects are studied under different names in different areas of mathematics and computer science (knots, curves, paths, traces, trajectories). In mathematics, 1-dimensional objects are well-understood, and research endeavors have moved on to higher dimensions. On the other hand, many fundamental applications demand solutions that deal with 1-dimensional objects, and these computational problems have largely been studied in separate communities by those unaware of all of the mathematical foundations.

The main goal of the proposed seminar was to identify connections and seed new research collaborations along the spectrum from knot theory and topology, to computational topology and computational geometry, all the way to graph drawing. Each of the invited speakers explored synergies in algorithms concerning 1-dimensional objects embedded in 2- and 3-dimensional spaces, as this is both the most fundamental setting in many applications, as well as the setting where the discrepancy in computational complexity between generic mathematical theory and potential algorithmic solutions is most apparent. In addition, each talk proposed a set of open questions from their research area that could benefit from attention from the other communities, and participants of the seminar were invited to propose their own research questions.

Below, we (the organizers) briefly describe the three main areas bridged; the abstracts of talks in the seminar and preliminary results from the working groups are also outlined later in this report.

Curves in Trajectory Analysis

Applications of computational topology are on the rise; examples include the analysis of GIS data, medical image analysis, graphics and image modeling, and many others. Despite how fundamental the question of topological equivalence is in mathematics, many of the relatively simple settings needed in computational settings (such as the plane or a 2-manifold) have been less examined in mathematics, where computability is known but optimizing algorithms in such "easy" settings has not been of interest until relatively recently.

Homotopy is one of the most fundamental problems to consider in a topological space, as this measure captures continuous deformation between objects. However, homotopy is notoriously difficult, as even deciding if two curves are homotopic is undecidable in a generic 2-complex. Nonetheless, many application settings provide restrictions that make computation more accessible. For example, most GIS applications return trajectories in a planar setting, at which point finding optimal homotopies (for some definition of optimal) becomes tractable.

Homology has been more recently pursued, as finding good homologies reduces to a linear algebra problem which can be solved efficiently. An example of this in the 1-dimensional setting is the recent work by Pokorny on clustering trajectories based on relative persistent homology. However, it is not always clear that optimal homologies provide as intuitive a notion for similarity measures compared with homotopy, and further investigations into applications settings is necessary.

Curves in Knot Theory

A fundamental question in 3-manifold topology is the problem of isotopy. Testing if two curves are ambiently isotopic is a foundational problem of *knot theory*: essentially, this asks whether two knots in 3-space are topologically equivalent. Problems in knot theory are tightly related to problems in 3-manifold topology, a field that has seen major breakthroughs in recent years, including Perelman's 2002 solution to the geometrisation and Poincaré conjectures, and Agol's recent 2012 proof of the virtual Haken conjecture. Algorithms and computation in these fields are now receiving significant attention from both mathematicians and computer scientists.

Complexity results are surprisingly difficult to come by. For example, one of the most fundamental and best-known problems is detecting whether a curve is knotted. This is known to be in both NP and co-NP; the former result was shown by Hass, Lagarias and Pippenger in 1999, but the latter was proven unconditionally by Lackenby just this year. Finding a polynomial time algorithm remains a major open problem. Hardness results are known for some knot invariants, but (despite being widely expected) no hardness result is known for the general problem of testing two knots for equivalence. Techniques such as randomisation and parameterised complexity are now emerging as fruitful methods for understanding the inherent difficulty of these problems at a deeper level.

Algorithmically, many fundamental problems in knot theory are solved by translating to 3-manifold topology. Here there have been great strides in practical software in recent years: software packages such as *SnapPy* and *Regina* are now extremely effective in practice for moderate-sized problems, and have become core tools in the mathematical research process. Nevertheless, their underlying algorithms have significant limitations: *SnapPy* is based on numerical methods that can lead to numerical instability, and *Regina* is based on polytope algorithms that can suffer from combinatorial explosions. It is now a major question as to how to design algorithms for knots and 3-manifolds that are exact, implementable, and have provably viable worst-case analyses.

Curves in Graph Drawing

On the computer science end of the spectrum, the study of one-dimensional objects is closely related to Graph Drawing.

Graph Drawing studies the embedding of zero- and one-dimensional features (vertices and edges of graphs) into higher-dimensional spaces; both from an analytic (given an embedding, what can we say about it) and synthetic (come up with a good embedding) point of view. Computational questions (how can we embed a given graph such that it satisfies certain properties / optimises certain criteria) and fundamental questions (which classes of graphs admit which styles of embeddings) have been studied extensively, and a large body of algorithmic results is readily available.

Planarity (non-crossing edges) is a central theme in graph drawing. There is a rich literature discussing which graphs can be drawn planarly, when, and how, as well as how to avoid crossings or other undesirable features in a drawing, such as non-rational vertices. Traditionally, edges have always been embedded as straight line segments; however, there is a recent trend to consider different shapes and curves, drastically increasing the space of possible drawings of a graph. The potential benefits of this broader spectrum are obvious, but the effects (both computational and fundamental) are still ill understood.

Connections between graph drawing and knot theory have long been recognised, yet are still being actively explored. Already in 1983, Conway and Gordon showed that every spatial representation of K_7 contains at least one knotted Hamiltonian cycle. Based on this, in 2013, Politano and Rowland characterised which knots appear as Hamiltonian cycles in canonical book embeddings of complete graphs (as defined by Otsuki in 1996).

Goals and Results of this Seminar

Now is an exciting time for computational and algorithmic knot theory: practical algorithms are showing their potential through experimentation and computer-assisted proofs, and we are now seeing key breakthroughs in our understanding of the complex relationships between knot theory and computability and complexity theory. Early interactions between mathematicians and computer scientists in these areas have proven extremely fruitful, and as these interactions deepen it is hoped that major unsolved problems in the field will come within reach.

Similarly, applications for graph drawing and trajectory analysis are in great demand, especially given the rise of massive amounts of data through GIS systems, map analysis, and many other application areas. However, despite the fact that many problems on curves are seen as mathematically trivial, there are few CS researchers who are truly familiar with the deeper topological results from mathematics. It is likely that many algorithmically interesting questions can benefit from an understanding of this rich history and toolset.

This seminar brought together a group of researchers from computer science and mathematics that study algorithms and mathematical properties of curves in various settings, as the interplay between these two groups is recent. In addition, we invited researchers in applications domains, who often do heuristic analysis of 1-dimensional objects in a variety of settings. Working groups were formed organically, but often allowed participants from various subfields to swap both open problems and favorite tools, and the overview talks discussed favorite tools and techniques from subdomains that may be useful to those in other areas. Concretely, we hope that in addition to the work begun in the working groups, many of these new collaborations will have positive long-term effects on all areas.

2 Table of Contents

3 Overview of Talks

3.1 Geometric Realizations and Reconfigurations

Anna Lubiw (University of Waterloo, CA)

Main reference Soroush Alamdari, Patrizio Angelini, Fidel Barrera-Cruz, Timothy M. Chan, Giordano Da Lozzo, Giuseppe Di Battista, Fabrizio Frati, Penny Haxell, Anna Lubiw, Maurizio Patrignani, Vincenzo Roselli, Sahil Singla, Bryan T. Wilkinson, "How to Morph Planar Graph Drawings", arXiv:1606.00425v1 [cs.CG], 2016.
URL https://arxiv.org/abs/1606.00425

Main results on drawing planar graphs deal with drawing edges as straight-line segments and restricting vertices to a small grid. I will discuss these issues for the problem of morphing (or "reconfiguring") one drawing of a planar graph to another. This can be done – while preserving a straight-line planar drawing – by means of a sequence of $O(n)$ linear morphs, where a linear morph moves each vertex at uniform speed along a straight line. ["How to Morph Planar Graph Drawings", to appear, SIAM J. Computing]. Restricting vertex positions (between the linear morphs) to a small grid is an open problem. Going beyond planarity to simultaneous planarity or intersection graphs of segments, we arrive at problems where finding a realization with straight line segments is complete for existential theory of the reals, and the reconfiguration space becomes disconnected.

3.2 Untangling Graphs and Curves on Surfaces via Local Moves

Hsien-Chih Chang (University of Illinois – Urbana-Champaign, US)

Joint work of Hsien-Chih Chang, Jeff Erickson
Main reference Hsien-Chih Chang, Jeff Erickson, "Untangling planar curves", in Proc. of the 32nd Int. Symp. on Computational Geometry (SoCG'16), LIPIcs, Vol. 51, pp. 29:1–29:15, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2016.
URL http://dx.doi.org/10.4230/LIPIcs.SoCG.2016.29

Any continuous deformation of one closed curve to another on the same surface can be decomposed into a finite sequence of local transformations called homotopy moves. We are interested in the number of homotopy moves required to simplify a generic closed curve with n self-crossings as much as possible on an arbitrary surface. In the plane, an $O(n^2)$ upper bound is implicit in the classical work of Steinitz on polyhedra; a later result of Hass and Scott extended this upper bound to contractible curves on arbitrary surfaces.

Electrical transformations – the collection of degree-1 reductions, series-parallel reductions, and ΔY transformations – was studied intensively due to its use in optimization problems on planar graphs. Again we are interested in the number of electrical transformations required to reduce a plane graph with n vertices as much as possible. Using arguments of Noble and Welsh, we can relate the number of electrical transformations required to reduce a plane graph to the number of homotopy moves required to simplify its medial graph, viewed as curves embedded in the plane. A major open problem due to Feo and Provan is whether $O(n^{3/2})$ electrical transformations are always sufficient.

In this talk we will survey the results on these two closely related problems, including the three classical approaches in the plane, the $\Theta(n^{3/2})$ bound on the number of homotopy moves required to simplify a plane curve, and a new result that simplifying a contractible curve in the annulus requires $\Omega(n^2)$ homotopy moves and its connection to the Feo and Provan conjecture.

This is a joint work with Jeff Erickson. Some of the results are published in our previous SoCG paper and its earlier preprint; the newer results can be found in our upcoming paper.

3.3 Embeddings in 3-Space

Eric Sedgwick (DePaul University – Chicago, US)

The author, along with Matuošek, Tancer and Wagner, showed that $\text{EMBED}_{2\to3}$, the problem of determining whether 2-complexes embed in 3-space is decidable. Here we discuss the obstacles, the intuition behind the solution, and the connection with Kirby diagrams, framed graphs embedded in 3-space that describe 3-manifolds. Finally, some open problems about embeddings and Kirby diagrams are stated.

3.4 Telling 3-manifolds apart: new algorithms to compute Turaev-Viro invariants

Jonathan Spreer (FU Berlin, DE)

Joint work of Benjamin Burton, Clément Maria, Jonathan Spreer
Main reference Clément Maria, Jonathan Spreer, "A polynomial time algorithm to compute quantum invariants of 3-manifolds with bounded first Betti number", in Proc. of the 28th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA'17), pp. 2721–2732, SIAM, 2017.
URL http://dx.doi.org/10.1137/1.9781611974782.180

In low-dimensional topology, distinguishing between manifolds is a fundamental problem, which is remarkably difficult to solve in dimensions beyond two. As a result, topologists rely on simpler invariants to solve this task. In dimension three, the Turaev-Viro invariants are amongst the most powerful invariants, but traditional algorithms to compute them have prohibitive running times for numerous instances occurring in practice.

In this talk I present a fixed parameter tractable algorithm to compute one of these invariants in polynomial time for manifolds with bounded Betti number. Moreover, I discuss further ideas and approaches for new algorithms and applications.

References

1 Clément Maria, Jonathan Spreer, *A polynomial time algorithm to compute quantum invariants of* 3*-manifolds with bounded first Betti number.* Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA 2017), pages 2721–2732, 2017.
2 Clément Maria, Jonathan Spreer, *Admissible colourings of 3-manifold triangulations for Turaev-Viro type invariants.* Proceedings of the 24th Annual European Symposium on Algorithms (ESA 2016), 64:1–64:16.
3 Benjamin A. Burton, Clément Maria, Jonathan Spreer, *Algorithms and complexity for Turaev-Viro invariants.* Automata, Languages, and Programming: 42nd International Colloquium (ICALP 2015), Proceedings, Part 1, pp. 281–293.

3.5 Persistent Cohomology and Circle-valued Coordinates

Mikael Vejdemo-Johansson (CUNY College of Staten Island, US)

We review the classical definition of simplicial homology and cohomology, demonstrate how the circle being the classifying space of $H^1(-;\mathbb{Z})$ produces equivalence classes of coordinate maps $[- \to S^1]$, and give examples from geometry, dynamics and motion capture time series.

3.6 Similarity Measures for Curves on Surfaces

Erin Moriarty Wolf Chambers (St. Louis University, US)

The question of how to measure similarity between curves in various settings has received much attention recently, motivated by applications in GIS data analysis, medical imaging, and computer graphics. While geometric measures such as the Hausdorff and Frechet distance have efficient algorithms, measures that take the underlying topology of the ambient space into account are less well understood. Several candidates have been proposed in recent years that are based on homotopy or homology, but many of these are only tractable in restricted settings, and surprisingly little is known about their practicality. In this talk, we will survey known results (both geometric and topological), and then focus on some of the recent algorithmic results and remaining open questions for the topological measures.

4 Working groups

4.1 Trajectory Clustering

Anne Driemel (TU Eindhoven, NL), Maike Buchin (Ruhr-Universität Bochum, DE), Brittany Terese Fasy (Montana State University – Bozeman, US), Florian T. Pokorny (KTH Royal Institute of Technology – Stockholm, SE), Mikael Vejdemo-Johansson (CUNY College of Staten Island, US), and Carola Wenk (Tulane University, US)

The starting point of our discussion was the intention to find a mathematically founded definition of clustering for curves which can be computed efficiently. Many clustering formulations are based on the definition of a centroid or median. This lead us to re-thinking the definition of a median for a set of curves. We observed that one can extend Tukey's definition of a median to our setting as follows. Recall that the Tukey depth of a point p in a finite set of points $P \subset R^d$ is defined as

$$\min_{\substack{h \in H \\ p \in h}} |P \cap h|,$$

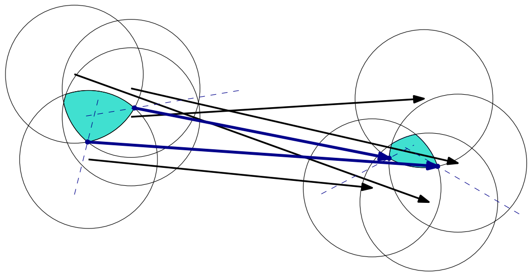


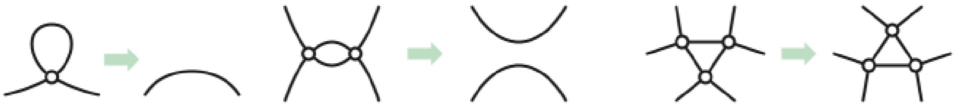
Figure 1 4 line segments (black) contained in a bisector represented by 2 line segments (blue).

where H is the set of half-spaces in R^d. The Tukey median is then defined as the point with largest depth. We observe that a half-space can be represented using the bisector of two points, which is the set of points that are equidistant to two fixed points a and b. This notion of half-space partition of P naturally extends to any distance metric defined on curves. To solve the above optimization problem we are thus interested in all 2-cell Voronoi partitions of P, where the Voronoi partitions are formed under a certain distance measure. We can represent a bisector by two curves a and b which we call *bisector representatives*. Given a bisector in this implicit way, we can compute the Voronoi partition by simply computing the distances of all points in P to the representatives a and b. In fact the count of points on both sides of the bisector can be estimated very efficiently by using random sampling on P. Most importantly, these computations can be done without computing the bisector explicitly. To solve the optimization problem we initially focused on the special case of line segments in R^2 and the Fréchet distance to measure distances between line segments. We think that one can compute all 2-cell Voronoi partitions by finding bisectors that contain subsets of points from P. In particular we believe that it is sufficient to consider subsets of either 4 or 5 line segments, and for each configuration it is sufficient to compute a constant number of bisector representatives. This directly implies a polynomial time algorithm for determining the median of a set of line segments. Furthermore we investigated geodesics in this space and how to project a curve onto its closest point on a given bisector.

4.2 Simplifying Curves on Surface via Local Moves

David Letscher (St. Louis University, US), Gregory R. Chambers (University of Chicago, US), Hsien-Chih Chang (University of Illinois – Urbana-Champaign, US), Arnaud de Mesmay (University of Grenoble, FR), Anne Driemel (TU Eindhoven, NL), Brittany Terese Fasy (Montana State University – Bozeman, US), Jessica S. Purcell (Monash University – Clayton, AU), Saul Schleimer (University of Warwick – Coventry, GB), Eric Sedgwick (DePaul University – Chicago, US), Dylan Thurston (Indiana University – Bloomington, US), Stephan Tillmann (University of Sydney, AU), and Birgit Vogtenhuber (TU Graz, AT)

A total of 12 seminar participants participated in discussions about the following question presented at the open problem session: “How many homotopy moves are needed to reduce a generic closed curve on a surface to have minimal number of self-intersections?"



Homotopy moves $1 \to 0$, $2 \to 0$, and $3 \to 3$.

Any closed curve on a surface can be *reduced* to another one that has minimum number of self-intersections within its homotopy class. Hass and Scott [5] showed that such curves can be reduced through a finite sequence of local transformations called *homotopy moves*. For generic curves in the plane with n self-intersections, a proof that $O(n^2)$ moves are always sufficient to reduce the curve is implicit in Steinitz's proof that every 3-connected planar graph is the 1-skeleton of a convex polyhedron [6, 7]. Specifically, Steinitz proved that any non-simple closed curve with no empty loops contains a minimal *bigon* which can be reduced and removed by a sequence of homotopy moves. This upper bound was later improved to $O(n^{3/2})$ by Chang and Erickson, which is the best possible in the worst case [1]. For curves on the annulus, de Graff and Schrijver [3] showed that $O(n^2)$ moves are always sufficient. Chang and Erickson [2] found a matching lower bound (which extends to curves in any higher genus surfaces).

Our group focused on finding upper bounds for the problem on various surfaces. We started with the torus and were able to show that $O(n^2)$ moves are again sufficient. Using similar techniques we were then able to prove the same quadratic upper bounds on the Möbius strip, the Klein bottle, and in the projective plane. In the final few meetings we considered the case of a curve with n self-intersections in a orientable surface of genus g. Using techniques from combinatorial group theory, we showed that there is an $O(g^2n^3)$ upper bound on number of homotopy moves required to reduce the given curve. Our main technical contribution is to extend Steinitz's bigon reduction technique to *singular bigons*, whose existence is guaranteed by an earlier result of Hass and Scott's [4].

A subset of participates of this group plans to write up the results and submit them to an appropriate conference or journal.

References

1 Hsien-Chih Chang and Jeff Erickson. Untangling planar curves. *Proc. 32nd Int. Symp. Comput. Geom.*, 29:1–29:15, 2016. Leibniz International Proceedings in Informatics 51.

2 Hsien-Chih Chang and Jeff Erickson. Unwinding annular curves and electrically reducing planar networks. Preprint, 2017.

3 Maurits de Graaf and Alexander Schrijver. Making curves minimally crossing by Reidemeister moves. *J. Comb. Theory Ser. B* 70(1):134–156, 1997.

4 Joel Hass and Peter Scott. Intersections of curves on surfaces. *Israel J. Math.* 51:90–120, 1985.

5 Joel Hass and Peter Scott. Shortening curves on surfaces. *Topology* 33(1):25–43, 1994.

6 Ernst Steinitz. Polyeder und Raumeinteilungen. *Enzyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* III.AB(12):1–139, 1916.

7 Ernst Steinitz and Hans Rademacher. *Vorlesungen über die Theorie der Polyeder: unter Einschluß der Elemente der Topologie*. Grundlehren der mathematischen Wissenschaften 41. Springer-Verlag, 1934. Reprinted 1976.

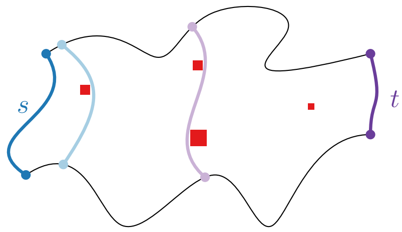


Figure 2 The Euclidean setting: two boundary paths (black), a start and end leash (s and t), and point obstacles (squares) with cost indicated by size. Two intermediate leashes are indicated.

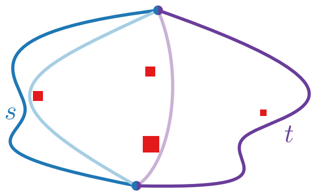


Figure 3 Special case with collapsed boundary paths.

4.3 Homotopy Height

Wouter Meulemans (TU Eindhoven, NL), Benjamin Burton (The University of Queensland, AU), Tim Ophelders (Eindhoven Univ. of Technology, NL), Bettina Speckmann (TU Eindhoven, NL), Marc van Kreveld (Utrecht University, NL), and Erin Moriarty Wolf Chambers (St. Louis University, US)

We initially considered the following problem, as posed in the open problem session: Given a triangulation of a disk, we wish to find the best way to move the left side of the boundary of the disk to the right side of the boundary, via flips across the faces or spikes along the edges. This problem is known as the *homotopy height problem*, as the sweep encodes a homotopy across the disk. One can also consider shrinking the boundary of the disk to a point, or the variant where one sweeps across an annulus, moving the outer boundary to the inner one step at a time.

In an effort to focus on a situation which allows for more geometric intuition, we studied the Euclidean case, in which we have a polygonal boundary enclosing a planar domain containing point obstacles (Figure 2). In this setting the boundary is divided into four parts: two *leashes* interleaved with two *boundary paths*. Our goal is now to continuously transform the starting leash s into the target leash t via a *homotopy*, in which the endpoints of the intermediate leashes travel along the boundary paths. The *cost* of a leash is its length plus an additional cost per obstacle it encounters; we refer to this additional cost as the *weight* of the obstacle. Our goal is to find a homotopy that minimizes the cost of the most expensive intermediate leash.

It follows from previous work that there is actually an optimal homotopy that is an *isotopy*. Furthermore, there is such an optimal isotopy in which the leashes move monotonically. We consider two different scenarios: all obstacles have the same (unit) weight, or obstacles can have different weights.

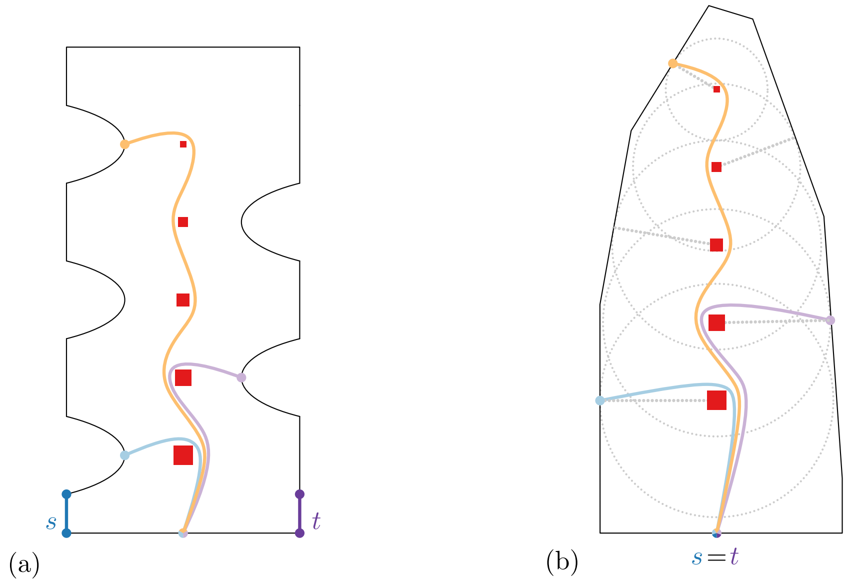


Figure 4 (a) A leash may need many inflection points in the variable-weight case. (b) The same principle in the convex case, with one boundary path and the two initial leashes collapsed to a point. Closest point and corresponding distance circles indicated for each obstacle.

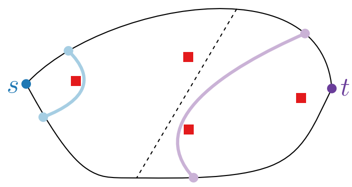


Figure 5 Special case with collapsed s and t, and unit-weight obstacles.

Variable-weight obstacles

If the two boundary paths collapse to points (Figure 3), then we can compute an optimal homotopy in polynomial time using a simple greedy strategy. Our result builds on the observation that in this setting the leashes do not have inflection points which are not induced by the boundary.

In the more general case, the leashes have no simple characterization anymore: specifically, any optimal homotopy might require a leash with linearly many inflection points. For this, consider the example in Figure 4(a). Here the weights are decreasing in the upward direction, to ensure that the best position of going over an obstacle is when the leash is as short as possible, only wrapping around the lower points. These best positions alternate between the first (left) half and the second (right) half of the boundary path, to create inflection points. This same principle can even be applied if the boundary is convex, only one boundary path is not a point, and both s and t are points (see Figure 4(b)). This leads us to conjecture that the problem is NP-Hard in this most general setting.

Unit-weight obstacles

For unit-weight obstacles we can compute an optimal homotopy in polynomial time, for the general case. Our results build on the same greedy strategy as in the variable-weight case.

For unit-weight obstacles, the leash does not need more than one inflection point that is not caused by the boundary, and this only in particular situations. Combining these observations with the monotonicity of an optimal homotopy allows us to solve this problem via dynamic programming.

We also found that some cases can be solved more efficiently. For example, if s and t collapse to a point (Figure 5), and the boundary paths together form a convex polygon, we need to find only an antipodal pair that splits the problem into two simpler ones: how do we shrink the leash at this antipodal pair to s (and analogously, to t)? This again follows the greedy strategy, and afterwards we only need to glue the two homotopies together, which may require at most one inflection point on the leash.

4.4 Convexifying Planar Drawings with Few Convexity-Increasing Linear Morphs

Lena Schlipf (FernUniversität in Hagen, DE), Linda Kleist (TU Berlin, DE), Boris Klemz (FU Berlin, DE), Irina Kostitsyna (Free University of Brussels, BE), Anna Lubiw (University of Waterloo, CA), Regina Rotman (University of Toronto, CA), Frank Staals (Aarhus University, DK), Darren Strash (Colgate University – Hamilton, US), and Carola Wenk (Tulane University, US)

Abstract

We study the problem of *convexifying* drawings of planar graphs. Given a planar straight-line drawing of a graph G, we wish to *morph* the drawing to a planar straight-line drawing of G in which all faces are convex, while maintaining planarity at all times. Furthermore, we want the morph to be *convexity-increasing*, meaning that the set of convex angles in the drawing never decreases. We give a polynomial time algorithm to construct such a morph.

Problem Definition and Background

A *morph* between two planar straight-line drawings Γ_0 and Γ_1 of a graph G is a continuous movement of the vertices from one to the other, with the straight-line edges determined by the vertex positions. If each vertex moves along a straight line at uniform speed, the morph is called *linear.* (Note that different vertices may move at different speeds, and some may remain stationary.) If, in addition, all the lines along which vertices move are parallel, then the morph is called *unidirectional.* A morph is *planar* if it preserves planarity of the drawing. Alamdari et al. [2] gave a polyomial time algorithm to find a planar morph of Γ_0 to Γ_1 using a sequence of $O(n)$ unidirectional morphs, where n is the number of vertices of the graph. One disadvantage of this algorithm is that vertices become almost coincident and there is no bound on the number of bits required for the vertex coordinates in the $O(n)$ intermediate drawings between the unidirectional morphs. Ideally, we would hope for the intermediate drawings to lie on a polynomial-sized grid (i.e. with a logarithmic number of bits for vertex coordinates). A weaker but still open question is to find planar morphs that can be specified with polynomially-bounded space (measuring bit complexity).

We say that a planar morph *convexifies* a given straight-line graph drawing if the end result is a convex graph drawing, i.e. a straight-line graph drawing in which the angles of all internal faces are convex and the angles of the external face are reflex. Note that we do not require strict convexity – we allow angles of 180°. We say that a planar morph is *convexity-increasing* if the set of face angles that are internal and convex, or external and reflex, never decreases, i.e. the progress towards a convex drawing, as measured by the number of face angles violating the convexity condition, is non-decreasing.

Throughout, we assume that our input graph has a convex drawing with the same faces and the same outer face as the input drawing. Necessary and sufficient conditions for the existence of such a convex drawing were given by Tutte [10], Thomassen [9], and Hong and Nagamochi [7]. These conditions can be tested in linear time by the algorithm of Chiba et al. [4]. Such conditions are usually stated for a fixed convex drawing of the outer face, but the conditions become simpler when, as in our case, the drawing of the outer face may be chosen to have no 3 consecutive collinear vertices. The conditions simplify further when no internal vertex has degree 2, and this can be assumed without loss of generality since an internal vertex of degree 2 must be drawn as a point in the interior of the straight line segment formed by its two incident edges. With these simplifications, the result stated by Hong and Nagamochi [7] is that a plane graph G with outer face C and with no internal degree-2 vertex has a convex drawing with outer face C if and only if the graph is *internally 3-connected*, i.e., the graph is 2-connected and any pair of cut vertices u, v has the properties that u and v lie on the outer face and every connected component of $G - \{u, v\}$ has a vertex of the outer face.

The algorithm of Alamdari et al. (or any other algorithm to morph graph drawings) can be used to convexify a given planar graph drawing Γ_0, since we can construct a convex drawing Γ_1 of the graph and morph Γ_0 to Γ_1. However, all known morphing algorithms triangulate the drawing, and hence will fail to be convexity-increasing in general[1]. Furthermore, morphing to *some* convex drawing is a weaker condition than morphing to a *particular* convex drawing, and may give us more freedom to keep to a small grid.

It is an open question to find convexity-increasing morphs. In the special case when the graph consists of a single cycle the problem is solved by the result of Aichholzer et al. [1] that morphs a polygon to a convex polygon without losing any visibilities between pairs of vertices.

Progress and Preliminary Results

At the seminar we outlined an algorithm to convexify a given planar graph drawing via a convexity-increasing morph that consists of $O(n)$ unidirectional morphs. Furthermore, each unidirectional morph moves vertices in either the horizontal or vertical direction, which means that the trajectory of each vertex during the complete morph is a path consisting of horizontal and vertical segments.

We will first discuss the reason why we concentrate on unidirectional morphs. After that we discuss the main idea of our algorithm.

Restricting to linear morphs seems like a sensible way to discretize morphs – essentially, it asks for the vertex trajectories to be piece-wise linear. At first glance, the restriction to unidirectional morphs, seems arbitrary and restrictive. However, it turns out to be easier to prove the existence of unidirectional morphs, for the following reason. Suppose we do a unidirectional morph in the direction parallel to the x-axis. Then every vertex must keep its

[1] We note that there is an algorithm to morph one convex drawing to another [3], which does not triangulate the graph, but this does not solve the problem of convexifying a non-convex drawing.

y-coordinate. This simplifies the planarity requirements because Lemma 13 of the paper by Alamdari et al. [2] states that the linear morph between two planar straight-line drawings Γ_1 and Γ_2 is planar if every line parallel to the x-axis crosses the same set of edges and vertices in the same order in both drawings. Note that this condition requires in particular that every vertex is at the same y-coordinate in both drawings, and the condition implies that the morph is unidirectional.

This means that, after committing to a direction for a unidirectional morph, we are free to choose a new drawing that satisfies the above conditions – and a planar unidirectional morph is guaranteed.

We use an existing algorithm to create new drawings. Planar straight-line drawings with vertices at fixed y-coordinates are called *layered drawings of hierarchical graphs*. Hong and Nagamochi [7] gave an algorithm to construct a convex layered drawing of any *hierarchical st-graph* – a hierarchical graph in which the boundary of every face consists of two upward chains.

Our algorithm proceeds by a sequence of steps where each step is as follows: choose the horizontal or vertical direction; augment the graph to a hierarchical *st*-graph; appeal to Hong and Nagamochi to produce a new convex layered drawing of this augmented graph; and perform a linear morph to the new drawing. A horizontal step will convexify any reflex angle formed by three vertices whose y-coordinates are increasing. Similarly, a vertical step will convexify any reflex angle formed by three vertices whose x-coordinates are increasing. No step will make a convex angle reflex. We may need to apply a shear transformation (which is a unidirectional morph) before each step in order to guarantee that there is at least one reflex angle whose vertices have increasing x- or y-coordinates.

Our main contribution is the following theorem.

▶ **Theorem 1.** *Any planar straight-line drawing of an internally* 3*-connected graph can be convexified via a convexity-increasing morph that consists of* $O(b)$ *unidirectional morphs, where* b *is the number of face angles that are internal and reflex or external and convex, and each unidirectional morph moves vertices in the horizontal or vertical direction. Furthermore, there is a polynomial-time algorithm to find such a morph.*

We have a family of examples, based on those of Alamdari et al. [2], to show that $\Omega(b)$ is a lower bound on the number of linear morphs that may be required to convexify a straight-line planar graph drawing.

Open Questions and Future Work

Although our algorithm improves on the general morphing algorithm [2] in that we do not use the technique of "almost" contracting vertices, still, we do not seem to have a polynomial bound on the bit-complexity of the intermediate drawings of our morph. We can design a family of examples, based on those of Lin and Eades [8] to show that one of our unidirectional morphs may unavoidably blow up the the width of the drawing from $O(n)$ to $\Omega(n!)$. This is still polynomial bit complexity, but the danger is that successive steps might cause exponential growth. In an attempt to get better bounds on the grid size/bit complexity of the intermediate drawings of the morph, we tried replacing Hong and Nagamochi's algorithm to find convex layered drawings by using an extension of Tutte's planar graph drawing algorithm [11] to more general edge weights ([5] or see [6]), and choosing the edge weights to place vertices at the desired layers. We plan to explore this issue further, but so far it remains an open question to achieve polynomially-bounded bit complexity.

Acknowledgments. We wish to thank André Schulz for helpful discussions on generalizations of Tutte's algorithm.

References

1 O. Aichholzer, G. Aloupis, E. D. Demaine, M. L. Demaine, V. Dujmovic, F. Hurtado, A. Lubiw, G. Rote, A. Schulz, D. L. Souvaine, and A. Winslow. Convexifying polygons without losing visibilities. In *CCCG*, 2011.
2 S. Alamdari, P. Angelini, F. Barrera-Cruz, T. M. Chan, G. Da Lozzo, G. Di Battista, F. Frati, P. Haxell, A. Lubiw, M. Patrignani, V. Roselli, S. Singla, and B. T. Wilkinson. How to morph planar graph drawings. *arXiv:1606.00425*, 2016. to appear in SIAM J. Computing.
3 P. Angelini, G. Da Lozzo, F. Frati, A. Lubiw, M. Patrignani, and V. Roselli. Optimal morphs of convex drawings. In L. Arge and J. Pach, editors, *Proceedings of the 31st International Symposium on Computational Geometry (SoCG 2015)*, volume 34 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 126–140, Dagstuhl, Germany, 2015. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
4 N. Chiba, K. Onoguchi, and T. Nishizeki. Drawing plane graphs nicely. *Acta Informatica*, 22(2):187–201, 1985.
5 M. S. Floater. Parametric tilings and scattered data approximation. *International Journal of Shape Modeling*, 4(03n04):165–182, 1998.
6 S. J. Gortler, C. Gotsman, and D. Thurston. Discrete one-forms on meshes and applications to 3D mesh parameterization. *Computer Aided Geometric Design*, 23(2):83–112, 2006.
7 S.-H. Hong and H. Nagamochi. Convex drawings of hierarchical planar graphs and clustered planar graphs. *Journal of Discrete Algorithms,* 8(3):282–295, 2010.
8 X. Lin and P. Eades. Towards area requirements for drawing hierarchically planar graphs. *Theoretical computer science*, 292(3):679–695, 2003.
9 C. Thomassen. Plane representations of graphs. In J. A. Bondy and U. S. R. Murty, editors, *Progress in Graph Theory*, pages 43–69. Academic Press, 1984.
10 W. T. Tutte. Convex representations of graphs. *Proceedings of the London Mathematical Society*, 3(1):304–320, 1960.
11 W. T. Tutte. How to draw a graph. *Proceedings of the London Mathematical Society*, 3(1):743–767, 1963.

4.5 The Pachner Graph of the Three-Sphere and Related Questions

Jonathan Spreer (FU Berlin, DE), Benjamin Burton (The University of Queensland, AU), Arnaud de Mesmay (University of Grenoble, FR), Parker Evans (Tulane University – New Orleans, US), Jessica S. Purcell (Monash University – Clayton, AU), and Saul Schleimer (University of Warwick – Coventry, GB)

Background

It is known that any two triangulations of a surface with the same number of vertices can be connected by a sequence of *diagonal flips*, or *2-2 moves*, given by exchanging the diagonal of a quadrilateral, as shown on the left of Figure 6. For example, see [1].

Similarly, any closed orientable 3-manifold can be decomposed into tetrahedra, and any two such decompositions with the same number of vertices can be related by a sequence of 3-dimensional flips, namely a *Pachner moves*, or 2-3 (3-2) moves, as shown on the right of

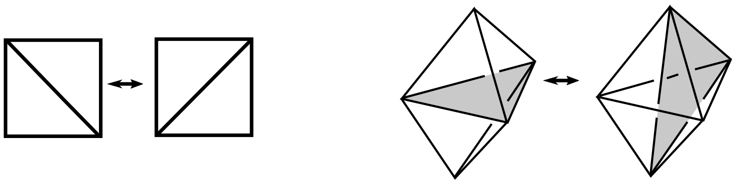


Figure 6 Left: A 2-dimensional diagonal flip. Right: A 3-dimensional flip, called a 2-3 move, or Pachner move. Two tetrahedra become three.

Figure 6. This is shown by Matveev in [6]. [**Note:** There are additional moves that are also called Pachner moves, such as the 4-1 and 1-4 moves, but these create and destroy vertices, so we ignore these for the purposes of this report.]

For a given 3-manifold and a given number of vertices, we build a graph related to these moves called the *Pachner graph*. Each vertex corresponds to a triangulation of the 3-manifold, and two vertices are connected by an edge if there is a single Pachner move changing one to the other. Matveev's result implies that the graph is connected.

Questions

The motivating major question is: what is the "shape" of the Pachner graph? Given two triangulations of a manifold with the same number of vertices, what is the shortest path between them? What is the shortest path to a "canonical" triangulation of the manifold?

These questions are hard, and wide open. For this project, we restrict to the 3-sphere S^3, and (typically) restrict to triangulations with a single vertex. Let T denote such a triangulation: a one-vertex triangulations of S^3, with n tetrahedra ($n > 1$).

Upper bounds

- There is a sequence of papers by Mijatovic (starting in 2003) which uses normal and almost normal surface theory to show that any triangulation of S^3 is connected to a constant size standard triangulation by a sequence of at most $A \cdot e^{Bn^2}$ moves [7, 8]. The bound comes from the upper bounds on the complexity of normal and almost normal spheres in T. Thus one might be able to find better upper bounds by finding smaller normal 2-spheres. However, there are examples of triangulations in which the smallest 2-sphere has exponentially large complexity (cf the examples of [5] and [4]).
- A natural tactic is to restrict to triangulations with nice properties. For example, the manifold could have a small Cheeger constant, or the dual graph to T could have low tree width. In this case, we can find a separator of small size. It is still unclear how to use such small separators to simplify triangulations.
- A final observation – much of the discussion here is similar to the story of Reidemeister simplification of diagrams of the unknot U. In that area there has been recent progress, due to Lackenby [3]. For any n crossing diagram of U there is a sequence of Reidemeister moves of length at most $O(n^{11})$ taking it to the trivial diagram. Can we use ideas similar to Lackenby's proof (using the combinatorics of a foliated spanning disk) to generalise to the 3-sphere? Or perhaps even the solid torus, where there is a simpler foliation?

So far we have the following possible correspondences:

Lackenby	Proposed 1	Proposed 2
unknot diagram	trig. of S^3	trig. of solid torus
unknot	1-skeleton	1-skeleton or μ
spanning disk	2-skeleton	2-skeleton or merid. disk
core curve	$\emptyset$	core curve
foliation by pages / book	foliation by spheres	foliation by annuli

Idea 1: Use Lackenby's short core curves [2] and annulus foliations to simplify triangulations of solid tori. [Aside: What is the algorithmic complexity of finding the short core curve?]

Idea 2: Take Lackenby's setup from [3] to prove Lackenby's theorem about core curves in [2], which Lackenby proves using other techniques.

Further questions:

1. **Unknotted edges:**
   - Does T having an unknotted edge help? Let T be a one-vertex triangulation of a 3-sphere with an unknotted edge. Is there a polynomial time algorithm to simplify T to a smaller triangulation? Is there a polynomial time algorithm to simplify T to a smaller triangulation T', where T' has an unknotted edge?
   - When does T have an unknotted edge? Ben Burton reports that every one-vertex triangulation of S^3, with at most nine tetrahedra, has an unknotted edge. The unknotted edge is not necessarily the highest valence edge (as we had expected).
   - Related: how do $2-3$ and $3-2$ moves change the complexity of the edges of T as knots in S^3 (for various notions of complexity)? Can we build a triangulation where edges are all knotted? Where they have arbitrarily high knot complexity (for various notions of complexity)?
2. **Tree width one:** Can we simplify 3-sphere triangulations with tree-width one in polynomial (quadratic) time?
   In the tree-width one case, Burton's thesis tells us that we either have a layered triangulation (which trivially can be simplified) or a "hat" (two triangles identified along two faces) with tetrahedra inside and outside. The latter complex can be simplified by a thickening-pulling-flattening move. Can we express this move in a sequence of Pachner moves?
3. **Other:**
   - Can we simplify locally constructible (collapsible) 3-sphere triangulations in polynomial time?
   - Let T be a 3-sphere with optimal Morse function with $\leq k$ critical faces. Can we simplify T in $O(f(k)n^{O(1)})$ ($O(f(k)e^{O(n)})$) time?

Lower bounds

Given a one vertex triangulation of S^3, can we find a lower bound on the number of moves required to simplify the triangulation? So far, the largest known examples take $n+2$ moves to simplify. Can we find something that requires more moves? We considered several families of triangulations of S^3.

1. Gluing a pair of Fibonacci layered tori, following ideas of Letscher [5]. These triangulations have several interesting properties, e.g. low tree-width, lots of very small separators, but *no* small two-spheres (even almost normal ones). However, these examples simplify directly (n moves). Interestingly there seems to be only one location for the simplification

to take place, namely between the layered tori. That is, there is some concentrated positive curvature.

2. Lackenby and Souto constructed "expansive spheres" [4]. These do not seem to appear in the literature, but we worked through some details of the construction, involving expansive Cayley graphs of the groups $\mathrm{PSL}(2, p^n)$. These examples do not have small tree-width and so do not have small separators of any kind. However, again these still simplify directly – there is concentrated positive curvature along the doubling sphere. This again seems to be the only place to do Pachner moves.
3. If a triangulation has only edges of degree five and higher, then there is no $3-2$-move possible, so we must increase the number of tetrahedra (at least twice!) before decreasing. For example the 600 cell has this property. [And leads to the minimal triangulation of the Poincaré homology sphere Σ^3.] Here the curvature seems to be well distributed. Of course, the triangulation has 120 vertices, not one. Note that Regina and Snappy both reduce this triangulation to bounded size (two tetrahedra) immediately.
   We did a search for one-vertex triangulations with only high valence edges, and found several with all edges of valence four and higher among the census with seven, eight, and nine tetrahedra. However, we found no triangulations of S^3 with edges of valence five and higher. We don't know of any obstruction to the existence of such triangulations.
4. Can we count the number of one-vertex, n-tetrahedra triangulations of S^3? Can we prove that the number of such triangulations grows super-exponentially? Does this imply lower bounds?

Other

1. Instead of studying the Pachner graph of 1-vertex 3-sphere triangulations, look at their vertex links and how they change. This will give us a subset of the vertex set of the Pachner graph of 2-sphere triangulations. How sparse is this subset? What can we say for the edges in this complex?
2. The flip graph of n-gon triangulations is isomorphic to the 1-skeleton of the associahedron. Is there a similar polytope (or related object) for the Pachner graph of S^3 (M^3, etc.)?

References

1 Allen Hatcher, *On triangulations of surfaces*, Topology Appl. **40** (1991), no. 2, 189–194. MR 1123262

2 Marc Lackenby, *Core curves of triangulated solid tori*, Trans. Amer. Math. Soc. **366** (2014), no. 11, 6027–6050. MR 3256192

3 Marc Lackenby, *A polynomial upper bound on Reidemeister moves*, Ann. of Math. (2) **182** (2015), no. 2, 491–564. MR 3418524

4 Marc Lackenby and Juan Souto, *Expander knots and metrics*, Unpublished.

5 David Letscher, *Fibonacci triangulations of s^3*, personal communication.

6 Sergei Matveev, *Algorithmic topology and classification of 3-manifolds*, second ed., Algorithms and Computation in Mathematics, vol. 9, Springer, Berlin, 2007. MR 2341532

7 Aleksandar Mijatović, *Simplifying triangulations of S^3*, Pacific J. Math. **208** (2003), no. 2, 291–324. MR 1971667

8 Aleksandar Mijatović, *Triangulations of Seifert fibred manifolds*, Math. Ann. **330** (2004), no. 2, 235–273. MR 2089425

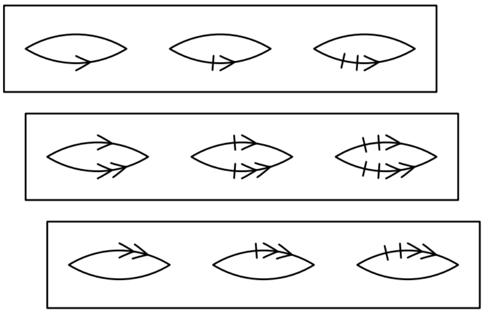


Figure 7 A parking garage, shown for $k = 3$ and $l = 3$. Edges with corresponding marking are glued; this can be achieved by stacking the sheets in 3 dimensions and attaching connecting ramps.

4.6 Quadratic Genus with Linear Boundary

Dylan Thurston (Indiana University – Bloomington, US), Benjamin Burton (The University of Queensland, AU), Maarten Löffler (Utrecht University, NL), Saul Schleimer (University of Warwick – Coventry, GB), and Stephan Tillmann (University of Sydney, AU)

In the course of analyzing multi-level motion planning [3, 1, 2], the following problem naturally arises.

▶ **Question 1.** *Suppose you are given a surface Σ smoothly embedded in $\mathbb{R}^3$ so that the vertical projection of Σ to $\mathbb{R}^2$ is an immersion with polygonal boundary made of m line segments. (That is, the vertical direction is transverse to Σ.) Is $\text{genus}(\Sigma)$ bounded by a linear function of m?*

We answered Question 1 in the negative; the best bound is quadratic. It turns out that the restriction to surfaces that are embedded in $\mathbb{R}^3$ is inessential.

▶ **Theorem 2.** *Let Σ be a surface with boundary and let $f : \Sigma \to \mathbb{R}^2$ be an immersion on the interior of Σ so that $f(\partial\Sigma)$ is a polygonal path with m line segments. Then $\text{genus}(\Sigma) \leq m(m+1)/4$. Furthermore, there are examples coming from embeddings in $\mathbb{R}^3$ with $\text{genus}(\Sigma) = (m/8 - 1)^2$.*

The examples achieving quadratic genus growth are "parking garages" $P_{k,l}$, as shown in Figure 7:

- take k parallel rectangular sheets;
- cut out l slits from each sheet (stacked on top of each other); and
- rejoin across the slits, shifting down one level as you go.

We can apply the Gauss–Bonnet theorem to the metric on Σ coming from the map to $\mathbb{R}^2$. We use the special case when the curvature vanishes on the interior and the curvature of the boundary is zero except at the polygonal corners.

▶ **Theorem 3** (Gauss–Bonnet, flat version). *Let Σ be a surface with a locally Euclidean metric and polygonal boundary, with corners at c_i with interior angle θ_i. Then the Euler characteristic of Σ is*

$$\chi(\Sigma) = 2 - 2\text{genus}(\Sigma) - \#\partial\Sigma = \frac{1}{2\pi}\sum_i(\pi - \theta_i),$$

where $\#\partial\Sigma$ is the number of components in the boundary of Σ.

Here, $\pi - \theta_i$ should be thought of as the bending angle at c_i: zero if there is no actual corner, positive if the corner is convex as on the boundary of a convex polygon in the plane, and negative if the corner is concave. Some of the corners in $P_{k,l}$ are very concave, with a total internal angle of approximately $2l\pi$. The result of this computation is that $\text{genus}(P_{k,l}) = (k-1)(l-1)$. Furthermore, $P_{k,l}$ can be realized with a polygonal boundary with $4k + 4l$ corners.

For the upper bounds on genus, we again apply Theorem 3 and give an upper bound on the interior angles θ_i. To do this, we first bound the total multiplicity in any region, the degree by which it is covered by Σ. The multiplicity at a point $x \in \mathbb{R}^2$ can be computed by sending a ray out to infinity in either direction from x, and so is at most $m/2$. The angle θ_i at a corner c_i is bounded by 2π times the multiplicity in any adjoining region. This yields the stated upper bound on genus.

References

1 A. Hillebrand, M. van den Akker, R. Geraerts, and H. Hoogeveen, *Performing multicut on walkable environments*, Combinatorial Optimization and Applications. COCOA 2016 (Cham) (T-H. Chan, M. Wang, and L. Wang, eds.), Lecture Notes in Computer Science, vol. 10043, Springer, 2016, pp. 311–325.

2 A. Hillebrand, M. van den Akker, R. Geraerts, and H. Hoogeveen, *Separating a walkable environment into layers*, Proceedings of the 9th International Conference on Motion in Games (New York, NY, USA), MIG'16, ACM, 2016, pp. 101–106.

3 W. van Toll, A. F. Cook, and R. Geraerts, *Navigation meshes for realistic multi-layered environments*, 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, Sept 2011, pp. 3526–3532.

4.7 Lombardi Drawings of Knots and Links

Birgit Vogtenhuber (TU Graz, AT), Philipp Kindermann (FernUniversität in Hagen, DE), Stephen G. Kobourov (University of Arizona – Tucson, US), Maarten Löffler (Utrecht University, NL), Martin Nöllenburg (TU Wien, AT), André Schulz (FernUniversität in Hagen, DE), and Dylan Thurston (Indiana University – Bloomington, US)

Introduction

This work is motivated by the following question posed by Benjamin Burton: Given a drawing of a knot, how can it be redrawn *nicely* without changing the given topology of the drawing?

A *knot* is an embedding of a simple closed curve in 3-dimensional Euclidean space R^3, considered up to continuous transformations, which cannot be untangled to the simple loop, also known as the *unknot*. Similarly, a *link* is a collection of simple closed curves in R^3 that cannot be untangled. A *drawing of a knot (link)* is a mapping of the knot (link) to the Euclidean plane R^2 such that for any point of R^2, at most two points of the curve(s) are mapped to it [6, 5, 1].

It is easy to see that drawings of links and knots are 4-regular plane multigraphs that contain neither loops nor split vertices. Likewise, every 4-regular planar multigraph without loops and split vertices can be interpreted as a link.

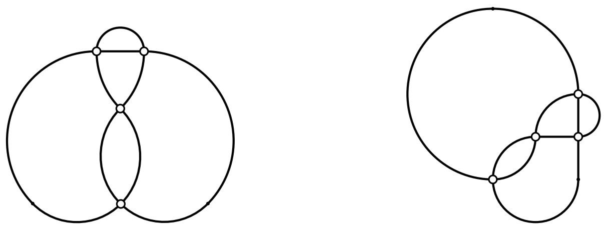


Figure 8 Two different 2-Lombardi drawings of knot 4_1, which by Theorem 2 does not admit a Lombardi-drawing.

A *Lombardi drawing* of a (multi-)graph $G = (V, E)$ is a drawing of G in the Euclidean plane with the following properties:

1. The vertices are represented as distinct points in the plane
2. The edges are represented as circular arcs connecting the representations of their end vertices (and not containing the representation of any other vertex); note that a straight-line segment is a valid circular arc with radius infinity.
3. Every vertex has *perfect angular resolution*, that is, its incident edges are equiangularly spaced. For links and knots this means that the angle between any two consecutive edges is $\pi/2$.

Lombardi drawings have been introduced by Duncan et al. [3] who showed a number of positive results (e.g., all d-regular graphs with $d \not\equiv 2 \pmod 4$ have circular Lombardi drawings and all 2-degenerate graphs have Lombardi drawings) and negative results (e.g., there are planar graphs that do not have planar Lombardi drawings). Eppstein [4] showed that every (simple) planar graph with maximum degree three has a plane Lombardi drawing. Further, he showed that a certain class of 4-regular planar graphs (the medial graphs of polyhedral graphs) also admit plane Lombardi drawings and presented an example of a 4-regular planar graph that does not have a plane Lombardi drawing.

k-Lombardi drawings are a generalization of Lombardi drawings in which every edge is a sequence of at most k circular arcs that meet at a common tangent. Duncan et al. [2] showed that every planar graph has a 3-Lombardi drawing.

Results

The main question we are considering in this work is motivated by applying the Lombardi drawing style to knot and link drawings: Given a 4-regular planar multigraph G with a fixed combinatorial embedding (without loops and split-vertices), does G admit a plane Lombardi drawing with the same embedding? And what can still be done if this is not the case?

By the results of Duncan et al. [2], every link trivially admits a 3-Lombardi drawing. As our first result, we showed that every link also admits a 2-Lombardi drawing.

▶ **Theorem 1.** *Every 4-regular planar multigraph G (without loops and split-vertices) with a fixed combinatorial embedding admits a plane 2-Lombardi drawing with the same embedding.*

Concerning the original Lombardi drawings, we know from Eppstein [4] that there exist 4-regular planar graphs that do not admit a Lombardi drawing. However, the example by Eppstein represents a link, not a knot. When searching for whether or not all knots admit a Lombardi drawing, we obtained a surprising negative result: The 4-knot is not Lombardi. Moreover, the following stronger statement holds:

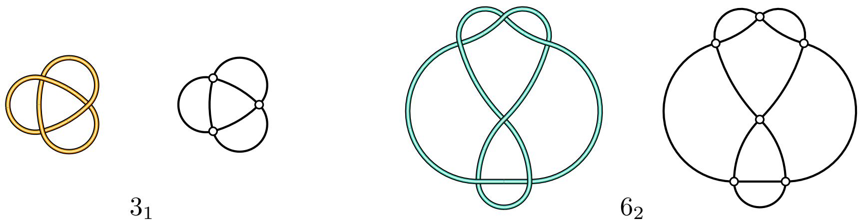


Figure 9 Two examples of Lombardi drawings of knots and the according 4-regular graphs whose existence follows from Theorem 3.

▶ **Theorem 2.** *Every 4-regular planar multigraph G that contains K_4 as a subdrawing does not admit a plane Lombardi drawing.*

On the positive side, we were able to extend the result from Eppstein [4] to a larger class of graphs: Every plane drawing D of a 4-regular planar multigraph can be interpreted as the medial graph of a multigraph and its dual. If one of those graphs is simple, then D admits a Lombardi-drawing.

▶ **Theorem 3.** *Let D be a 4-regular planar multigraph G (without loops and split-vertices) with a fixed combinatorial embedding and let M and M' be the primal-dual multigraph pair for which D is the medial graph. If one of M and M' is simple, then D admits a plane Lombardi drawing with the same embedding.*

We remark that neither the result from Theorem 2 nor the one from Theorem 3 is tight: We found a 4-regular planar multigraph G that does not contain the 4-knot as a subdrawing and still does not admit a Lombardi drawing, and we found 4-regular planar multigraphs admitting a Lombardi drawing whose primal-dual pair M and M' both contain parallel edges.

Open problems and ongoing work

There are many open questions remaining which we plan to consider in this context. As main questions concerning Lombardi drawings we have the following.

▶ **Question 4.** *Can we give a complete characterization of 4-regular planar multigraphs that admit a Lombardi drawing?*

▶ **Question 5.** *What is the complexity of deciding whether a given 4-regular planar multigraph admits a Lombardi drawing?*

The next question is about the transition between Lombardi drawings and 2-Lombardi drawings.

▶ **Question 6.** *Given a 4-regular planar multigraph, what is the minimum number of edges consisting of two circular arcs in any 2-Lombardi drawing?*

We conclude with a question about a different relaxation of Lombardi drawings for drawings of 4-regular planar multigraphs.

▶ **Question 7.** *Can we redraw every drawing of a 4-regular planar multigraph using circular arcs as edges such that at every vertex, every pair of non-adjacent edges emanates in opposite directions? And if yes, what is the maximum smallest angle between consecutive edges that we can guarantee?*

References

1 P. R. Cromwell. *Knots and links*. Cambridge University Press, 2004.

2 C. A. Duncan, D. Eppstein, M. T. Goodrich, S. G. Kobourov, and M. Löffler. Planar and poly-arc Lombardi drawings. In M. van Kreveld and B. Speckmann, editors, *Graph Drawing (GD'11)*, volume 7034 of *LNCS*, pages 308–319. Springer, 2012.

3 C. A. Duncan, D. Eppstein, M. T. Goodrich, S. G. Kobourov, and M. Nöllenburg. Lombardi drawings of graphs. *Journal of Graph Algorithms and Applications*, 16(1):37–83, 2012.

4 D. Eppstein. Planar Lombardi drawings for subcubic graphs. In W. Didimo and M. Patrignani, editors, *GD 2012*, number 7704 in LNCS, pages 126–137. Springer-Verlag Berlin Heidelberg, 2013.

5 D. Rolfsen. *Knots and links*. American Mathematical Soc., 1976.

6 R. G. Scharein. *Interactive Topological Drawing*. PhD thesis, Department of Computer Science, The University of British Columbia, 1998.

Participants

- Maike Buchin
Ruhr-Universität Bochum, DE
- Benjamin Burton
The University of Queensland, AU
- Gregory R. Chambers
University of Chicago, US
- Hsien-Chih Chang
University of Illinois – Urbana-Champaign, US
- Arnaud de Mesmay
University of Grenoble, FR
- Anne Driemel
TU Eindhoven, NL
- Parker Evans
Tulane University – New Orleans, US
- Brittany Terese Fasy
Montana State University – Bozeman, US
- Philipp Kindermann
FernUniversität in Hagen, DE
- Linda Kleist
TU Berlin, DE
- Boris Klemz
FU Berlin, DE
- Stephen G. Kobourov
University of Arizona – Tucson, US
- Irina Kostitsyna
Free University of Brussels, BE
- David Letscher
St. Louis University, US
- Maarten Löffler
Utrecht University, NL
- Anna Lubiw
University of Waterloo, CA
- Wouter Meulemans
TU Eindhoven, NL
- Martin Nöllenburg
TU Wien, AT
- Tim Ophelders
Eindhoven University of Technology, NL
- Florian T. Pokorny
KTH Royal Institute of Technology – Stockholm, SE
- Jessica S. Purcell
Monash University – Clayton, AU
- Regina Rotman
University of Toronto, CA
- Saul Schleimer
University of Warwick – Coventry, GB
- Lena Schlipf
FernUniversität in Hagen, DE
- André Schulz
FernUniversität in Hagen, DE
- Eric Sedgwick
DePaul University – Chicago, US
- Bettina Speckmann
TU Eindhoven, NL
- Jonathan Spreer
FU Berlin, DE
- Frank Staals
Aarhus University, DK
- Darren Strash
Colgate University – Hamilton, US
- Dylan Thurston
Indiana University – Bloomington, US
- Stephan Tillmann
University of Sydney, AU
- Marc van Kreveld
Utrecht University, NL
- Mikael Vejdemo-Johansson
CUNY College of Staten Island, US
- Birgit Vogtenhuber
TU Graz, AT
- Carola Wenk
Tulane University, US
- Erin Moriarty Wolf Chambers
St. Louis University, US

Report from Dagstuhl Seminar 17081

Computability Theory

Edited by

Klaus Ambos-Spies[1], Vasco Brattka[2], Rodney Downey[3], and Steffen Lempp[4]

1 **Universität Heidelberg, DE,** ambos@math.uni-heidelberg.de
2 **Universität der Bundeswehr – München, DE,** vasco.brattka@unibw.de
3 **Victoria University – Wellington, NZ,** rod.downey@vuw.ac.nz
4 **University of Wisconsin – Madison, US,** lempp@math.wisc.edu

Abstract

Computability is one of the fundamental notions of mathematics and computer science, trying to capture the effective content of mathematics and its applications. Computability Theory explores the frontiers and limits of effectiveness and algorithmic methods. It has its origins in Gödel's Incompleteness Theorems and the formalization of computability by Turing and others, which later led to the emergence of computer science as we know it today. Computability Theory is strongly connected to other areas of mathematics and theoretical computer science. The core of this theory is the analysis of relative computability and the induced degrees of unsolvability; its applications are mainly to Kolmogorov complexity and randomness as well as mathematical logic, analysis and algebra. Current research in computability theory stresses these applications and focuses on algorithmic randomness, computable analysis, computable model theory, and reverse mathematics (proof theory). Recent advances in these research directions have revealed some deep interactions not only among these areas but also with the core parts of computability theory. The goal of this Dagstuhl Seminar is to bring together researchers from all parts of computability theory and related areas in order to discuss advances in the individual areas and the interactions among those.

Seminar February 19–24, 2017 – http://www.dagstuhl.de/17081
1998 ACM Subject Classification F.1.1 Models of Computation, F.4.1 Mathematical Logic
Keywords and phrases algorithmic randomness, computability theory, computable algebra, computable analysis, generic case complexity, proof mining
Digital Object Identifier 10.4230/DagRep.7.2.89
Edited in cooperation with Rupert Hölzl

1 Executive Summary

Klaus Ambos-Spies
Vasco Brattka
Rodney Downey
Steffen Lempp

Computability theory grew from work to understand effectiveness in mathematics. Sophisticated tools have been developed towards this task. For a while, the area tended to be concerned with internal considerations such as the structure of the various hierarchies developed for the tasks of calibrations. More recently, this machinery has seen modern applications into areas such as model theory, algorithmic randomness, analysis, ergodic theory, number theory and the like; and the tools have been used to answer several classical questions. Seminar 17081 was an opportunity for researchers in several areas of modern computability theory to get together and interact.

The format was for 2–3 lectures in the morning with at least one being an overview, and a similar number of 3–4 in the afternoon, with Wednesday afternoon and Friday afternoon free. The weather being miserable, participants opted to stay at the Schloss Wednesday afternoon, and quite a bit of work was done in pairs in the time left free, on the Wednesday afternoon in particular. At least one problem seen as significant in the area was solved (one concerning the strength of Ramsey's Theorem for Pairs in reverse mathematics), and the organizers know of several other papers in preparation resulting from the seminar.

The lectures were from various areas, but the greatest concentration was around

- classification tools in computable analysis (the Weihrauch Lattice) and Reverse Mathematics (on what proof-theoretic strength is needed to establish results in mathematics), and these areas' relationship with generating algorithms, such as in proof mining;
- computable model theory (looking at structures such as groups, rings, or abstract algebraic structures, endowing them with effectivity and seeing what else is algorithmic). Notable was the new work on effective uncountable structures such as uncountable free groups, and on topological groups;
- algorithmic randomness: Here one seeks to give meaning to randomness for individual strings and infinite sequences. Talks given explored the relationship of calibrations of randomness to computational power, and online notions of randomness.

Of course, these are not separate areas but are inter-related, and the talks reflected these inter-relationships.

Currently, computability theory is quite vibrant with many new applications being found, and a number of young and talented researchers entering the field. This was reflected in the age of the presenters of many of the lectures, as well as the significant number of people we could have invited in addition.

All in all, the meeting was a great success and should have significant impact on the development of the area.

2 Table of Contents

3 Overview of Talks

3.1 Machines running on random tapes and the probabilities of events

George Barmpalias (Victoria University – Wellington, NZ)

Joint work of Andrew Lewis-Pye, George Barmpalias, Douglas Cenzer, Christopher P. Porter
URL http://barmpalias.net

Probabilistic Turing machines have been studied since the 1940s, when it was shown that the probability of a machine with a random (as in probability theory) oracle computing any fixed non-computable real is 0. Chaitin's halting probability is the probability that a universal Turing machine halts on a random oracle (with empty numerical input) and was characterized in terms of algorithmic randomness and computable approximations. In general, one can ask the same question with respect to any property that a computation of a universal Turing machine may have when it reads a random oracle:

1. Will it compute a total function?
2. Will it enumerate a co-finite set (say, as the domain of a partial function that it computes)?
3. Will it enumerate a set which computes the halting problem?
4. Will it compute an incomputable function?
5. Will it halt with an output inside a certain set A?

Can we give characterizations of these probabilities in terms of algorithmic randomness and effectiveness properties? We show that this is possible, but we do not always get the expected answer.

Moreover we answer one of the last remaining questions from the BSL 2006 list of open questions in randomness (by Miller and Nies), by showing that the probability that the universal machine halts and outputs a number in a non-empty Π^0_1 set is always left-c.e. and ML-random. Intuitively, this says that if we code arithmetical sentences into numbers, the probability that the universal machine outputs an undecidable sentence (in PA) can be effectively approximated from below!

My talk is mainly based on the following recent work:

- The probability of a computable output from a random oracle
  George Barmpalias, Douglas Cenzer and Christopher P. Porter
  https://arxiv.org/abs/1612.08537
- Differences of halting probabilities
  George Barmpalias and Andy Lewis-Pye
  https://arxiv.org/abs/1604.00216
- Random numbers as probabilities of machine behaviour
  George Barmpalias, Douglas Cenzer and Christopher P. Porter
  https://arxiv.org/abs/1605.05838

3.2 Deep Π^0_1 classes

Laurent Bienvenu (University of Montpellier & CNRS, FR)

Joint work of Laurent Bienvenu, Christopher P. Porter
Main reference L. Bienvenu, C. P. Porter, "Deep Π^0_1 classes", Bulletin of Symbolic Logic, 22(2):249–286, 2016.
URL https://arxiv.org/abs/1403.0450v3

We will present the concept of deep Π^0_1 classes, which can be thought of as those classes whose paths are uniformly 'hard to generate probabilistically' and discuss the many interesting properties those classes enjoy. In particular we will see that they behave quite similarly to the class of PA degrees in their interactions with algorithmic randomness.

3.3 Finding bases of uncountable free abelian groups is hard

Noam Greenberg (Victoria University – Wellington, NZ)

Joint work of Noam Greenberg, Dan Turetsky, Linda Brown Westrick

We use admissible computability to discuss effective properties of uncountable free abelian groups. Assuming $V = L$, for all regular uncountable κ there is a κ-computable free abelian group with no κ-computable basis, indeed no κ-arithmetical basis, and usually one can avoid any lower cone below a $\Delta^1_1(L_\kappa)$ degree. On the other hand, not much can be coded into bases of groups: a forcing construction shows that the most that can be coded is $\emptyset'$ or $\emptyset''$, depending on κ (for example, if it is a successor of a singular cardinal, or inaccessible). The index-set of κ-computable free abelian groups is $\Sigma^1_1(L_\kappa)$-complete, unless κ is weakly compact.

3.4 Monte Carlo Computability

Rupert Hölzl (Universität der Bundeswehr – München, DE)

Joint work of Vasco Brattka, Rupert Hölzl, Rutger Kuyper
Main reference V. Brattka, R. Hölzl, R. Kuyper, "Monte Carlo Computability", Proc. of the 34th Symp. on Theoretical Aspects of Computer Science (STACS 2017), LIPIcs, Vol. 66, pp. 17:1–17:14, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017.
URL http://dx.doi.org/10.4230/LIPIcs.STACS.2017.17

We introduce Monte Carlo computability as a probabilistic concept of computability on infinite objects and prove that Monte Carlo computable functions are closed under composition. We then mutually separate the following classes of functions from each other: the class of multi-valued functions that are non-deterministically computable, that of Las Vegas computable functions, and that of Monte Carlo computable functions. We give natural examples of computational problems witnessing these separations. As a specific problem which is Monte Carlo computable but neither Las Vegas computable nor non-deterministically computable, we study the problem of sorting infinite sequences.

3.5 Strong and non-strong degrees of categoricity

Iskander Kalimullin (Kazan Federal University – Kazan, RU)

Joint work of Nikolay Bazhenov, Iskander Kalimullin, Mars Yamaleev

A computable structure $\mathcal{A}$ has a *degree of categoricity* $\mathbf{x}$ if $\mathbf{x}$ is the least Turing degree such that $\mathcal{A}$ is $\mathbf{x}$-computably categorical. A degree of categoricity $\mathbf{x}$ is *strong* if there are two computable copies $\mathcal{B} \cong \mathcal{C} \cong \mathcal{A}$ such that $\mathbf{x} \leq_{\mathrm{T}} f$ for every isomorphism $f : \mathcal{B} \to \mathcal{C}$. Answering a question from [1] on the existence of non-strong degrees of categoricity we introduce the notion of spectral dimension of a computable structure: the *spectral dimension* of a computable structure $\mathcal{A}$ with a degree of categoricity $\mathbf{x}$ is equal to an ordinal $n \leq \omega$ if n is the least ordinal such that there are computable copies $\mathcal{B}_i \cong \mathcal{C}_i \cong \mathcal{A}$, $i < n$, such that $\mathbf{x} \leq_{\mathrm{T}} \bigoplus_{i<n} f_i$ for every choice of isomorphisms $f_i : \mathcal{B}_i \to \mathcal{C}_i, i < n$ (considering categoricity spectra the notion of spectral dimension can be easily adapted to the case when a structure has no degree of categoricity). We show that for every $n < \omega$ there is a rigid computable structure of the degree of categoricity $\mathbf{0}'$ having spectral dimension n. The original question from [1] now can be updated to the form: is there a computable structure with a degree of categoricity having spectral dimension ω? Such a structure, if it exists, can not be rigid.

References

1 E. B. Fokina, I. Kalimullin, R. Miller, "Degrees of categoricity of computable structures", Archive for Mathematical Logic, Vol. 49, pp. 51–67, 2010.

3.6 Topological aspects of enumeration degrees

Takayuki Kihara (University of California – Berkeley, US)

Joint work of Takayuki Kihara, Steffen Lempp, Keng Meng Ng, Arno Pauly

Pauly and the speaker introduced a general way of assigning a degree structure to each admissibly represented space. From this perspective, the enumeration degrees can be thought of as the degree structure of a universal second-countable T_0 space. This idea enable us to classify enumeration degrees in terms of general topology. For instance, the Turing degrees (total e-degrees) are the "finite dimensional metrizable e-degrees", and the continuous degrees are the "metrizable e-degrees". We can then talk about the existences of a Hausdorff (T_2) e-degree which is not an Urysohn ($T_{2.5}$) e-degree, of a Frechet (T_1) e-degree which is Hausdorff-quasimininal, etc.

Note that the admissibly represented spaces form a cartesian closed category, which is far larger than the category of second-countable T_0 spaces. In general, the degree structure of a non-second-countable space is not a substructure of the enumeration degrees. For instance, one can show that the de Groot dual of the Baire space (which is not second-countable) has a point having non-enumeration degree.

3.7 The Scott Isomorphism Theorem

Julia Knight (University of Notre Dame, US)

Joint work of Rachael Alvir, Julia Knight, Charles McCoy

Scott [5] proved that for a countable structure $\mathcal{A}$ for a countable language L, there is a sentence of $L_{\omega_1\omega}$ (a *Scott sentence*) whose countable models are just the isomorphic copies of $\mathcal{A}$. The complexity of an optimal *Scott sentence* measures the internal complexity of the structure. I will describe some recent results on the complexity of Scott sentences. I had conjectured that every finitely generated group has a d-Σ_2 Scott sentence, and every computable finitely generated group has a computable d-Σ_2 Scott sentence. Recently, Harrison-Trainor and Ho [2] showed that both conjectures are false. Alvir, McCoy and I [1] applied a result of Montalbán [4] and one of A. Miller [3] to show that a finitely generated group has a d-Σ_2 Scott sentence iff there is a generating tuple whose orbit is defined by a Π_1 formula. Using effective versions of the results of Montalbán and A. Miller, we get the fact that a computable finitely generated group has a computable d-Σ_2 Scott sentence iff there is a generating tuple whose orbit is defined by a computable Π_1 formula.

References

1 R. Alvir, J. F. Knight, C. McCoy, "Complexity of Scott sentences", Preprint.
2 M. Harrison-Trainor, M-C. Ho, "Finitely generated groups", Preprint.
3 A. Miller, "The Borel classification of the isomorphism class of a countable model", Notre Dame Journal of Formal Logic, Vol. 24, pp. 22–34, 1983.
4 A. Montalbán, "A robuster Scott rank", Proceedings of the AMS, Vol. 143, pp. 5427-5436, 2015.
5 D. Scott, "Logic with denumerably long formulas and finite strings of quantifiers", In The Theory of Models, ed. by J. Addison, L. Henkin, and A. Tarski, pp. 329–341, North-Holland, 1965.

3.8 Computability, Proof Mining and Metric Regularity

Ulrich Kohlenbach (TU Darmstadt, DE)

Concepts of metric regularity and weak sharp minima which are generalizations of quantitative notions of strong uniqueness to problems with non-unique solutions play an important role in convex optimization. We will discuss computability and proof theoretic aspects of this as well as applications to minimization problems, fixed points of resolvents and zeros of subdifferentials (partly joint work with Genaro Lopez-Acedo). We also present recent applications of 'proof mining' to convex feasibility problems [2, 3]. In particular, we give a polynomial rate of asymptotic regularity [4] for Bauschke's solution of the minimal displacement conjecture [1], that is, for Picard iterates of compositions of metric projections in Hilbert space (without the assumption of the existence of a fixed point).

References

1 H. Bauschke, "The composition of projections onto closed sets in Hilbert space is asymptotically regular", Proceedings of the AMS, Vol. 131, pp. 141–146, 2003.

2 M. A. A. Khan, U. Kohlenbach, "Quantitative image recovery theorems", Nonlinear Analysis, Vol. 106, pp. 138–150, 2014.
3 U. Kohlenbach, "On the quantitative asymptotic behavior of strongly nonexpansive mappings in Banach and geodesic spaces", Israel Journal of Mathematics, Vol. 216, pp. 215–246, 2016.
4 U. Kohlenbach, "A polynomial rate of asymptotic regularity for compositions of projections in Hilbert space", Submitted.

3.9 A peek at the higher levels of the Weihrauch hierarchy

Alberto Marcone (University of Udine, IT)

Joint work of Andrea Cettolo, Alberto Marcone

Weihrauch reducibility and the ensuing Weihrauch hierarchy have been successfully used to refine reverse mathematics results for statements which are provable in ACA_0 and below. The study the Weihrauch hierarchy for functions arising from statements lying at higher levels (such as ATR_0) of the reverse mathematics spectrum was suggested by the author in the open problem session of Dagstuhl Seminar 15392 in September 2015.

We start this study, obtaining in some cases the expected finer classification, but in others observing a collapse of statements that are not equivalent with respect to provability in subsystems of second order arithmetic. This is in part due to the increased syntactic complexity of the statements.

Our preliminary results deal with comparability of well-orderings, $\mathbf{\Sigma}^1_1$-separation, and $\mathbf{\Delta}^1_1$-comprehension.

3.10 Randomness notions in Muchnik and Medvedev degrees

Kenshi Miyabe (Meiji University – Kawasaki, JP)

The main question of this talk is whether one can construct a more random set from a random set. This question can be formalized by mass problems, that is, Muchnik and Medvedev degrees. As an example, computable randomness is strictly below ML-randomness in Muchnik degrees because there exists a high minimal Turing degree, which contains a computably random set but no ML-random set is Turing below it. Similar arguments can be applied for other pairs. There are two interesting pairs of randomness notions that have the same Muchnik degree. One pair is the one of ML-randomness and difference random. This is because, for ML-random set $X + Y$, at least one of X or Y should be difference random. In contrast, ML-randomness and difference random have different Medvedev degrees. In other words, one can not compute a difference random from a ML-random uniformly. The proof uses the Levin-Kautz theorem and no-randomness-from-nothing for ML-randomness. The other pair is the one of Schnorr randomness and computable randomness. They have the same Muchnik degree but different Medvedev degrees. The proof extends the method separating Schnorr randomness and computable randomness using Levy's zero-one law from probability theory.

3.11 Stopping time complexity

Alexander Shen (University of Montpellier & CNRS, FR)

Joint work of Mikhail Andreev, Gleb Posobin, Alexander Shen

Consider a bit string x written on the input one-directional tape of some Turing machine. We want the machine to stop reading the tape exactly when x is read. How much information should be communicated to this machine? We may call this amount "stopping time complexity" of x.

This quantity (in the context of prediction theory) was considered by Vovk and Pavlovic (see https://arxiv.org/abs/1603.04283), and we try to perform a more systematic analysis of it in the language of Kolmogorov complexity.

One can consider the plain version of stopping time complexity (minimal plain complexity of a Turing machine that stops at x). It turns out to be equivalent to monotone-conditional complexity $C(x|x*)$ where the condition x is considered as a prefix of the string. There is also a quantitative characterization as a minimal upper semicomputable function such that on every path there is at most 2^n points where the function drops below n.

We show also that one should be careful: for the general case of $C(x|y*)$ we should consider monotone (prefix-stable), not prefix-free functions of y.

A similar theory can be constructed for prefix versions of stopping time complexity. We answer the question asked by Vovk–Pavlovic and show that the minimal prefix complexity of a program stopping at x, the quantity $\mathrm{K}(x|x*)$ and the logarithm of stopping time semimeasure, introduced by Vovk and Pavlovic, are all different. Also we show that the stopping time semimeasure has a natural probabilistic interpretation while for the general case $m(x|y*)$ the natural interpretation is no longer valid.

3.12 Genericity, randomness, and differentiable functions

Sebastiaan A. Terwijn (Radboud University Nijmegen, NL)

Joint work of Rutger Kuyper, Sebastiaan A. Terwijn
Main reference R. Kuyper, S. A. Terwijn, "Effective genericity and differentiability", Journal of Logic and Analysis, 6(4):1–14, 2014.
URL http://logicandanalysis.org/index.php/jla/article/view/215/94

We present a theorem characterizing the notion of 1-genericity in terms of differentiable functions. We compare this to a recent characterization of the notion of 1-randomness, also in terms of differentiability. We also discuss variations about n-genericity, multiple differentiability, and polynomial time computability.

References

1 V. Brattka, J. S. Miller, A. Nies, "Randomness and Differentiability", Transactions of the American Mathematical Society, Vol. 368, pp. 581–605, 2016.
2 R. Kuyper, S. A. Terwijn, "Effective genericity and differentiability", Journal of Logic and Analysis, Vol. 6(4), pp. 1–14, 2014.

3.13 Stochasticity in Algorithmic Statistics for Polynomial Time

Nikolay K. Vereshchagin (Moscow State University, RU & National Research University Higher School of Economics – Moscow, RU) and Alexey Milovanov

Main reference A. Milovanov, N. Vereshchagin, "Stochasticity in Algorithmic Statistics for Polynomial Time", Report TR17-043, ECCC, 2017.
URL https://eccc.weizmann.ac.il/report/2017/043/download

A fundamental notion in Algorithmic Statistics is that of a stochastic object, that is, an object having a simple plausible explanation. Informally, a probability distribution is a plausible explanation for x if it looks likely that x was drawn at random with respect to that distribution. In this paper, we suggest three definitions of a plausible statistical hypothesis for Algorithmic Statistics with polynomial time bounds, which are called *acceptability, plausibility and optimality.* Roughly speaking, a probability distribution μ is called an acceptable explanation for x, if x possesses all properties decidable by short programs in a short time and shared by almost all objects (with respect to μ). Plausibility is a similar notion, however this time we require x to possess all properties T decidable even by long programs in a short time and shared by almost all objects. To compensate the increase in program length, we strengthen the notion of 'almost all' – the longer the program recognizing the property is, the more objects must share the property. Finally, a probability distribution μ is called an optimal explanation for x if $\mu(x)$ is large (close to $2^{-C^{\text{poly}}(x)}$).

Almost all our results hold under some plausible complexity theoretic assumptions. Our main result states that for acceptability and plausibility there are infinitely many non-stochastic objects, that is, objects that do not have simple plausible (acceptable) explanations. We explain why we need assumptions – our main result implies that P $\neq$ PSPACE. In the proof of that result, we use the notion of an *elusive set*, which is interesting in its own right. Using elusive sets, we show that the distinguishing complexity of a string x can be super-logarithmically less than the conditional complexity of x with condition r for almost all r (for polynomial time bounded programs). Such a gap was known before, however only in the case when both complexities are conditional, or both complexities are unconditional.

It follows from the definition that plausibility implies acceptability and optimality. We show that there are objects that have simple acceptable but implausible and non-optimal explanations. We prove that for strings whose distinguishing complexity is close to Kolmogorov complexity (with polynomial time bounds) plausibility is equivalent to optimality for all simple distributions, a fact that can be considered a justification of the Maximal Likelihood Estimator.

3.14 Turing-, tt-, and m-reductions for functions in the Baire hierarchy

Linda Brown Westrick (University of Connecticut – Storrs, US)

Joint work of Adam Day, Rod Downey, Linda Brown Westrick

For arbitrary functions $f\colon [0,1] \to \mathbb{R}$, (including in particular highly non-continuous functions), what would be the right notion of Turing reducibility and its variants? We present a computationally motivated definition of $\leq_{\mathrm{T}}$, $\leq_{\mathrm{tt}}$, and $\leq_{\mathrm{m}}$ for such functions, and show that within the Baire hierarchy, the linearly ordered $\leq_{\mathrm{T}}$-equivalence classes correspond precisely to the proper Baire classes. Further, within the Baire 1 functions, the $\leq_{\mathrm{tt}}$ and $\leq_{\mathrm{m}}$ equivalence classes enjoy a natural correspondence with levels of the α rank on Baire 1 functions considered in Kechris and Louveau (1990).

4 Working groups

4.1 Summary of the open problems session

- **Question (Barmpalias)** What is the measure of minimal covers in the Turing degrees?
- **Conjecture (Barmpalias)** A K-degree has uncountably many predecessors if and only if it is not infinitely often K-trivial.

- **Question (Fouché)** The continuous action of a topological group G on a discrete set X is said to be a Ramsey action if for each finite subset F of X and each finite colouring of X, there is some $g \in G$ such that the colouring is monochromatic on gF. Such an action is necessarily transitive. A topological group is called Ramsey if all transitive actions on discrete sets are Ramsey. Let LO be the set of total orders on the natural numbers, viewed as a closed subspace of $\mathbb{N}^{\mathbb{N}\times\mathbb{N}}$. It is a deep fact that if G is a Ramsey group, then the logical action of G on LO has a fixed point. It is well-known that the space LO has a unique S_∞-invariant Radon measure μ. This is a computable measure. The problem proposed is to understand the fixed points of the action of a Ramsey group on LO from the viewpoint of algorithmic randomness relative to μ.

- **Question (Kalimullin; see Abstract 3.5)** Is there a computable structure with a degree of categoricity having the spectral dimension ω?
- **Question (Kalimullin)** Is there a computable structure $\mathcal{A}$ of computable dimension 2 with 2-element automorphism group such that two isomorphisms between its computable copies $\mathcal{A}_0$ and $\mathcal{A}_1$ have incomparable Turing degrees?

- **Question (Nies)** For K-trivial sets A and B we say that $A \leq_{\mathrm{ML}} B$ if every ML-random A computing B computes A. Is $\leq_{\mathrm{ML}}$ arithmetical? Note that by the Gandy basis theorem, if $A \not\leq_{\mathrm{ML}} B$ then there is a counterexample $Z \leq_{\mathrm{T}} \mathcal{O}$.
- **Question (Nies)** A K-trivial set A is called smart if every ML-random $Y \geq_{\mathrm{T}} A$ computes all the K-trivials. Is there a minimal pair of smart K-trivials?
- **Question (Nies)** Suppose A is K-trivial. Is A Turing below each LR-hard ML-random?
- **Question (Nies)** Is weak 2-randomness closed upward under $\leq_{\mathrm{K}}$?

- **Question (Yu)** For any real x and constant c, let $A_{x,c} = \{n \mid \mathrm{K}^x(n) \geq \mathrm{K}(n) - c\}$. Define $x \geq_{\mathrm{WLK}} y$ if for any c, if $A_{x,c}$ is infinite, then there is some d so that $A_{x,c} \subseteq A_{y,d}$. The question is whether for any weakly low for K real x, $x \geq_{\mathrm{WLK}} y$ implies $x \geq_{\mathrm{LK}} y$?
- **Question (Yu)** Is it true that under the assumption of PD, every uncountable Π^1_3 set A ranges over an upper cone of Q_3-degrees?

5 Impact

5.1 Preliminary results as reported by participants

During the seminar, numerous groups of researchers took advantage of the free time between talks to collaborate on their current research projects. While at the time of writing of this report it is too early to fully appreciate the impact of the seminar on the field of computability, the following very partial list describes concrete results already reported by participants.

- Willem Fouché reports that he and Arno Pauly finalised two papers during the seminar: "How constructive is constructing measures?" (Journal of Logic and Analysis, 9:c3, 1–30, 2017) and "Weihrauch-completeness for layerwise computability" (together with George Davie; submitted to Logical Methods in Computer Science). He furthermore reports that he and André Nies continued their work on the project "Computable profinite groups and randomness".
- Ulrich Kohlenbach reports that he finished his work on the article "A polynomial rate of asymptotic regularity for compositions of projections in Hilbert space" (submitted March 2017) during the seminar.
- Antonio Montalbán and Richard Shore report that they worked briefly on finishing their article "Conservativity and ultrafilters over subsystems of second order arithmetic" which is about to be submitted. They also worked extensively on a follow-up article tentatively titled "Iterated Hindman's Theorem, Gower's Fin_k Theorem, and the Infinite Hale-Jewett Theorem all peas in a pod".
- Linda Brown Westrick is reporting that after her presentation on the topic of a continuous reducibility for Borel functions she had fruitful discussions with Takayuki Kihara, Antonio Montalbán, and Arno Pauly, who shared with her some connections with their work. She also took advantage of the workshop to work on an ongoing project about Scott sets with Mariya Soskova, and on another ongoing project about a lightface version of reducibility for Borel functions with Rod Downey.

Participants

- Klaus Ambos-Spies
Universität Heidelberg, DE
- George Barmpalias
Victoria University – Wellington, NZ
- Laurent Bienvenu
University of Montpellier & CNRS, FR
- Vasco Brattka
Universität der Bundeswehr – München, DE
- Peter Cholak
University of Notre Dame, US
- Chitat Chong
National University of Singapore, SG
- Barbara Csima
University of Waterloo, CA
- Rodney Downey
Victoria University – Wellington, NZ
- Willem L. Fouché
UNISA – Pretoria, ZA
- Sergey Goncharov
Sobolev Inst. of Mathematics – Novosibirsk, RU
- Noam Greenberg
Victoria University – Wellington, NZ
- Rupert Hölzl
Universität der Bundeswehr – München, DE
- Iskander Shagitovich Kalimullin
Kazan Federal University, RU
- Takayuki Kihara
University of California – Berkeley, US
- Julia Knight
University of Notre Dame, US
- Ulrich Kohlenbach
TU Darmstadt, DE
- Steffen Lempp
University of Wisconsin – Madison, US
- Alberto Marcone
University of Udine, IT
- Alexander Melnikov
Massey University, NZ
- Wolfgang Merkle
Universität Heidelberg, DE
- Joseph S. Miller
University of Wisconsin – Madison, US
- Russell G. Miller
CUNY Queens College – New York, US
- Kenshi Miyabe
Meiji University – Kawasaki, JP
- Antonio Montalbán
University of California – Berkeley, US
- Keng Meng Ng
Nanyang TU – Singapore, SG
- André Otfrid Nies
University of Auckland, NZ
- Arno Pauly
Free University of Brussels, BE
- Alexander Shen
University of Montpellier & CNRS, FR
- Richard A. Shore
Cornell University, US
- Theodore A. Slaman
University of California – Berkeley, US
- Andrea Sorbi
University of Siena, IT
- Mariya I. Soskova
University of Sofia, BG
- Sebastiaan A. Terwijn
Radboud University Nijmegen, NL
- Henry Towsner
University of Pennsylvania – Philadelphia, US
- Daniel Turetsky
Victoria University – Wellington, NZ
- Nikolay K. Vereshchagin
Moscow State University, RU & National Research University Higher School of Economics – Moscow, RU
- Wei Wang
Sun Yat-sen University – Guangzhou, CN
- Klaus Weihrauch
FernUniversität in Hagen, DE
- Linda Brown Westrick
University of Connecticut – Storrs, US
- Guohua Wu
Nanyang TU – Singapore, SG
- Yue Yang
National University of Singapore, SG
- Liang Yu
Nanjing University, CN

Report from Dagstuhl Seminar 17082

Shape-Changing Interfaces

Edited by

Jason Alexander[1], Sean Follmer[2], Kasper Hornbæk[3], and Anne Roudaut[4]

1 **Lancaster University, GB,** `j.alexander@lancaster.ac.uk`
2 **Stanford University, US,** `sfollmer@stanford.edu`
3 **University of Copenhagen, DK,** `kash@diku.dk`
4 **University of Bristol, GB,** `anne.roudaut@bristol.ac.uk`

Abstract

Shape-changing interfaces use physical shape change as input and output; such interfaces are emerging as an alternative way of interacting with computers. This seminar brought together researchers working on shape-changing interfaces to discuss three key themes: (1) The technologies involved in shape-change, including soft and modular robotics, smart materials, and mechanical actuation. (2) The design of shape-changing interfaces, including their key application areas, and their industrial and interaction design. (3) The user experience of shape-changing interfaces, including evaluations of such interfaces and psycho-physical evaluation results. The seminar set out to strengthen this new community, create opportunities for active collaborations, and to reach-out to other fields.

The seminar was attended by 25 researchers from around the world. These researchers represented the disciplines of Computer Science, Design, Engineering, Robotics and Material Science. This seminar had no formal presentations, but instead focused on working-group discussion and report-back sessions. This report outlines the key findings of these sessions.

Seminar February 19–22, 2017 – http://www.dagstuhl.de/17082
1998 ACM Subject Classification H.5.3. User Interfaces: Graphical User Interfaces, Input devices and strategies, Interaction Styles; H.5.m. Information interfaces and presentation (e.g. HCI): Miscellaneous.
Keywords and phrases shape-changing interfaces, user interfaces, materials, smart materials, robotics
Digital Object Identifier 10.4230/DagRep.7.2.102

1 Executive Summary

Jason Alexander
Sean Follmer
Kasper Hornbæk
Anne Roudaut

The Shape-Changing Interfaces Dagstuhl seminar aimed to bring together researchers from the disciplines of Computer Science, Design, Engineering, Robotics and Material Science to strenghten this new communinity, discuss grand challenges, form a research agenda, and to create opportunities for active collaborations.

Shape-Changing Interfaces, *Dagstuhl Reports*, Vol. 7, Issue 2, pp. 102–108
Editors: Jason Alexander, Sean Follmer, Kasper Hornbæk, and Anne Roudaut
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Shape-changing interfaces use changes in physical geometry to convey input and output and are emerging as an alternative interaction method for communicating with computers. Discussions at the seminar were based around three key themes: (1) The technologies involved in shape-change, including soft and modular robotics, smart materials, and mechanical actuation. (2) The design of shape-changing interfaces, including their key application areas, and their industrial and interaction design. (3) The user experience of shape-changing interfaces, including evaluations of such interfaces and psycho-physical evaluation results.

To encourage active discussion, the seminar had no keynote speakers, but instead used brainstorming activities and small working-groups to understand challenges, explore the literature, and plan an agenda. Specifically, the following sessions were run:

Benefits and Applications of Shape-Change: A whole-group brainstorming session developed categories of benefits and potential application areas for shape-changing interfaces.

Related Work: Small working-groups focused on one of five related-work areas (materials, hardware, experience and interaction, design, or applications), researched, and then presented summaries of the five ground-breaking and five most over-looked works in that sub-field.

Grand Challenges: A whole-group brainstorming session generated ideas and themes of grand challenges, small working groups then took a theme and dug deeper into the challenge, generating avenues of work and research agendas.

Worst Case Scenarios: To understand why this field could fail, a brainstorming exercise asked participants to develop a series of 'failure' situations—these were used as a method of creating awareness of the reasons progress in this field could stall.

Personal Reflections: To conclude the seminar, four participants were asked to provide their personal reflections on the experience, and their key take-home messages.

The seminar was attended by 25 researchers from around the world; all of who found the experience invaluable. This report outlines the key findings of these sessions.

2 Table of Contents

3 Seminar Sessions

Jason Alexander (Lancaster University, GB), Sean Follmer (Stanford University, US), Kasper Hornbæk (University of Copenhagen, DK), and Anne Roudaut (University of Bristol, GB)

Joint work of All seminar participants.

3.1 Benefits and Applications of Shape-Change

While participants at the seminar universally agreed on the benefits and potential of this field, articulating these benefits and potential applications provided a communal base for understanding everyone's perspectives and setting a research agenda. Participants were asked to brainstorm ideas in this area. These ideas were themed, falling into the categories of: accessibility, adaptability and reusability, collaboration and communication, portability, and returning the physicality and tactility of the real-world to computer interfaces. A large range of application areas and specific ideas were also generated.

3.2 Related Work

Seminar participants were then tasked with summarising the key related work in a particular sub-theme of shape-changing interfaces, with the organisers asking to see four groundbreaking papers and the four most over-looked articles in that sub-theme. Presentations were provided on Materials, Hardware, Experience and Interaction, Design, and Applications. Briefly, these reported on:

Materials: a range of novel materials and approaches for the implementation of shape-change devices, including work from Soft Robotics, Mechanical Meta-materials, inflatable materials, and epidermal electronics.

Hardware: the difference between between internally-powered and externally-powered shape-change, self-assembly and programmable matter, pin-arrays, mid-air/untethered shape-change, and self-actuated surfaces.

Experience and Interaction: progression from the Ultimate Display to Relief and InForm, taxonomies of shape-change, stiffness, and the under-representation of public installations.

Design: a tour of Stewart Platforms, the Computational Design of Mechanical Characters, thin-film based design, toy-inspired scenarios, and inspiration from commercial future visions.

Applications: scoped a wide range of applications already in the literature, including gaming, communication, notifications, collaborations, understanding data, sport, healthcare, rehabilitation, accessibility, and assisted living.

3.3 Grand Challenges

In this session, participants were asked to brainstorm and categorise grand-challenges in the development and deployment of shape-changing interfaces. The key challenge themes that emerged were: affordance and signifiers, arts and aesthetics, collaboration, economy, environmental impact, ethics and law, end-users, society, technology, and theory. Three of these themes were then chosen for deeper exploration:

Theory: this group identified many issues around theory in this space including: what is the purpose of theory for shape-change, there are many descriptive theories but nothing for future exploration, there is no theory for how we perceive motion, can we take theory from other disciplines, the unknown quantity of 'good' shape-change, what does the lack of theory stop us doing?

Cross-disciplinary Collaboration: there are significant issues to be tackled during cross-disciplinary collaboration. These include applications and visibility, cost efficiency, cultural challenges, ethics, geographic distance, language/terminology, publication (the 'what is a result?' question, author order), methods, proof/evaluation, and time-scales.

Affordance and Signifiers: this group studied the question: how do we communicate the fact that an object is shape-changing? We can consider the computer-as-tool, the computer-as-partner, and the computer-as-medium. We can use mapping, visibility, transfer effects, and feed-forward to illustrate the presence of shape-change.

3.4 Worst Case Scenarios

To better understand why and how the shape-changing interfaces field could fail, participants were asked to brainstorm their 'worst case scenario' for research and progression in this space. The resulting brainstorming input was classified into a number of themes: economic infeasibility, ethical and legal issues, lack of applications, other disciplines (e.g. robotics) achieve better solutions, never becomes socially acceptable, physicality becomes unnecessary, technology limitations, safety and trust, sustainability issues.

Despite the 'negative' approach of this activity, awareness of potential risks helps to enable mitigation strategies and reduce their chance of occurrence.

4 Personal Reflections

Panos Markopoulos (TU Eindhoven, NL), Pierre Dragicevic (INRIA, FR), Marianne Graves Petersen (Aarhus University, DK), and Isabel Qamar (University of Bristol, GB)

To conclude the seminar we asked four participants to provide their personal summary and review of the seminar and the key messages they would take away.

Professor Panos Markopoulos: Panos' review noted that we did not try to agree on a definition of shape-changing interface (a positive thing) and challenged us to consider how we should go beyond 'interfaces', suggesting the field could instead focus on 'physically interactive things'. The articulation of the large number of challenges faced by the field provides an excellent agenda for future research.

Dr. Pierre Dragicevic: Pierre's review encouraged participants to consider the differences between shape-changing objects and interfaces and to consider why shape-change is necessary. Further, understanding what is 'lost' by including shape-change in interfaces is just as important as what is gained. He ended with the question of what the new 'C' should be in HCI.

Dr. Isabel Qamar: Isabel underlined the importance of collaboration across disciplines in order to make this field successful. She raise the important question of how we disseminate research results during these collaborations, with different research fields valuing different contributions. As a material scientist, Isabel emphasised the importance of communicating the 'H' in HCI: what the humans need and want from other disciplines.

Dr. Marianne Graves Petersen: Marianne reflected on the scale of issues that were discussed during the seminar, noting we touched on small implementation issues through to large, meta-questions about economy and ethics that will ultimately dictate the success of shape-changing interfaces.

5 Conclusion and Next Steps

Jason Alexander (Lancaster University, GB), Sean Follmer (Stanford University, US), Kasper Hornbæk (University of Copenhagen, DK), and Anne Roudaut (University of Bristol, GB)

Joint work of All seminar participants.

This Dagstuhl Seminar was found extremely valuable by all attending researchers and it acted as an excellent conduit to strengthen the community around this exciting and innovative topic. As a result of this seminar we have created an active mailing list and are planning to create trimestrial email newsletter to keep the community strongly connected. We are also planning to develop a roadmap document using the output from this seminar to provide focus and direction for accelerating research in this area. All attendees expressed a desire to return to Dagstuhl in 3–5years to review progress and developments in this fast-moving field.

Participants

- Jason Alexander
Lancaster University, GB
- Miguel Bruns Alonso
TU Eindhoven, NL
- Céline Coutrix
Universität Stuttgart, DE
- Pierre Dragicevic
INRIA Saclay – Orsay, FR
- Aluna Everitt
Lancaster University, GB
- Sean Follmer
Stanford University, US
- Audrey Girouard
Carleton University – Ottawa, CA
- Marianne Graves Petersen
Aarhus University, DK
- Fabian Hemmert
Universität Wuppertal, DE
- Trevor Hogan
Cork Institute of Technology, IE
- Kasper Hornbaek
University of Copenhagen, DK
- Yvonne Jansen
UPMC – Paris, FR
- Matt Jones
Swansea University, GB
- Hyunyoung Kim
Universität Stuttgart, DE
- Majken Kirkegaard Rasmussen
Aarhus University, DK
- Panos Markopoulos
TU Eindhoven, NL
- Diego Martinez Plasencia
University of Sussex – Brighton, GB
- Timothy Merritt
Aarhus University, DK
- Isabel Qamar
University of Bristol, GB
- Anne Roudaut
University of Bristol, GB
- David Schmelzeisen
RWTH Aachen, DE
- Jürgen Steimle
Universität des Saarlandes, DE
- Sri Subramanian
University of Sussex – Brighton, GB
- Roel Vertegaal
Queen's University – Kingston, CA
- Helge A. Wurdemann
University College London, GB

Report from Dagstuhl Seminar 17091

Computer Science Meets Ecology

Edited by

Gustau Camps-Valls[1], Thomas Hickler[2], and Birgitta König-Ries[3]

1 **Universitat de València, ES,** gustau.camps@uv.es
2 **Senckenberg Research Centre, DE,** thomas.hickler@senckenberg.de
3 **Universität Jena & German Centre for Integrative Biodiversity Research (iDiv), DE,** birgitta.koenig-ries@uni-jena.de

Abstract

This report summarizes the program and main outcomes of the Dagstuhl Seminar 17091 entitled "Computer Science Meets Ecology". Ecology is a discipline that poses many challenging problems involving big data collection, provenance and integration, as well as difficulties in data analysis, prediction and understanding. All these issues are precisely the arena where computer science is concerned. The seminar motivation was rooted in the belief that ecology could largely benefit from modern computer science. The seminar attracted scientists from both fields who discussed important topics in ecology (e.g. botany, animal science, biogeochemistry) and how to approach them with machine learning, computer vision, pattern recognition and data mining. A set of specific problems and techniques were treated, and the main building blocks were set up. The important topics of education, outreach, data and models accessibility were also touched upon. The seminar proposed a distinctive perspective by promoting cross-fertilization in a unique environment and a unique set of individuals.

Seminar February 26 to March 3, 2017 – http://www.dagstuhl.de/17091
1998 ACM Subject Classification D.2.12 Interoperability, J.3 Life and Medical Sciences – Biology and genetics, I.6.5 Model Validation and Analysis
Keywords and phrases ecology, biodiversity, earth observation, earth system, remote sensing, computer science, citizen science, big data, data integration, modeling, semantics, society
Digital Object Identifier 10.4230/DagRep.7.2.109
Edited in cooperation with Ivaylo Kostadinov

1 Executive Summary

Gustau Camps-Valls
Joachim Denzler
Thomas Hickler
Birgitta König-Ries
Markus Reichstein

Ecology is a discipline that shows clearly the potential but also the challenges of computer supported research described as the 4th scientific paradigm by Jim Gray. It is increasingly data driven, yet suffers from hurdles in data collection, quality assurance, provenance, integration, and analysis.

We believe that ecology could profit from modern computer science methods to overcome these hurdles. However, usually, scientists in ecology are not completely aware of current trends and new techniques in computer science that can support their daily work. Such support could consist in the management, integration, and (semi-)automatic analysis of resources, like experimental data, images, measurements, in the generation of useful metadata, cloud computing, distributed processing, etc. Ecoinformatics is regarded as an important supporting discipline by many ecologists. However, up to now, very few computer scientists are involved in this discipline; mostly ecoinformatics (or biodiversity informatics) is done by people with a strong background in e.g. ecology and a long (mostly self-taught) experience in data management. It lacks a strong connection to cutting-edge computer science research in order to profit from the results of this area. On the other hand, computer scientists know too little about the domain to be able to offer solutions to relevant problems and to identify potential research avenues.

Motivated by our belief that a stronger bond between the disciplines that goes beyond viewing computer science as a "service provider" is of vital importance, we proposed this Dagstuhl seminar. The aim of the Dagstuhl seminar was to establish such links between (geo-)ecologists, ecoinformaticians and computer scientists.

The seminar: perspective and self-evaluation

Before the seminar. It turned out that it was not an easy task to motivate non-computer scientists to attend the seminar. For many, travel costs were a hurdle ultimately preventing attendance. This resulted in an unusually large number of declined invitations (often accompanied by "I would love to attend, but…" emails.

Despite these initial problems, we believe that the aim to start building links among the communities was reached at the seminar: We had fruitful discussions in numerous working groups resulting in some very concrete plans for future work.

Organization of the seminar. A total of 27 attendees gathered at the seminar. The wide variety of expertise and backgrounds constituted an initial challenge for the organization. The agenda considered a first round of presentations of the individuals and their research groups with a clear outline and items to treat (personal background, Research Areas/Interests, prospective links to „Computer Science meets Ecology“ seminar). After this, the main topics of interest for a wide audience were designed: essentially, three breakout groups were set up in the very first day of the meeting. Over the course of the seminar, these groups were adjusted, split up, or merged, several times. This resulted in quite a number of topics being touched upon with concrete results ranging from a working example for the application of a new method to a modeling problem to concrete plans for publications, a proposal and follow-up activities. Reports on these groups were given in the plenary session, and can be found in this report.

Broad results of the seminar. Results from the seminar can be categorized in three types: (i) collaborative and networking, as new joint works on specific topics came out of the meeting; (ii) knowledge transfer between fields, as computer scientists learned about the main problems in ecology involving data, while ecologists became aware of what kind of problems data scientists can solve nowadays; and (iii) educational, as several young PhD students and postdocs attended and participated in high level discussions.

Conclusions. The seminar brought together top scientists in the fields of ecology and computer science. The group of individuals was largely interdisciplinary, with a wide range of interests and expertises in each community too: from botany and animal science, to machine learning and computer vision. The seminar was organized in two main types of sessions: plenary and working group sessions to better focus on particular topics. Interesting developments and discussions took place in both, and a high level of cross-fertilization and future collaborations was initiated. On top of this, there was a broad consensus among the participants that the seminar should be the start of a series of yearly or bi-yearly meetings. We hope that the success of this first seminar will encourage broader participation in follow-up activities.

2 Table of Contents

3 Introduction

3.1 Why Computer Science needs to meet Ecology

Gustau Camps-Valls (University of Valencia, ES), Benjamin Adams (University of Auckland, NZ), Joachim Denzler (Universität Jena, DE), Thomas Hickler (Senckenberg Research Centre, DE), Birgitta König-Ries (Universität Jena, DE), Markus Reichstein (MPI für Biogeochemistry – Jena, DE), and Johann Wolfgang Wägele (ZFMK – Bonn, DE)

In his pioneering work, Jim Gray identified the 4th scientific paradigm, arguing that modern science needs computer supported research. Recent developments in many scientific disciplines prove him right: Huge amounts of heterogeneous, unstructured and multisource data can now be collected routinely, sometimes in a fully automatic manner. Due to the development of computer hardware and sensors even new data modalities are readily available. The main difference to the general "big data" hype is that in science collecting data always has the intention to gain insights into processes and mechanisms, or in general to gain knowledge from data, typically motivated by some hypothesis. So far, the main challenge is to manage the explosive growth in size, complexity, and rates of data accumulation. On the one hand, it is easy to collect terabytes of data per minute. On the other hand, analysing even a fraction out of it still remains a big problem for scientists, companies and international organizations. A discipline that shows the potential but also the challenges of this 4th scientific paradigm is Ecology.

Ecology is the study of the interactions amongst organisms and with their physical environment. For a long time, ecological analyses have been realized locally both with respect to both the geographical and phenomenological area of investigation. Today, scientists are interested in quantifying ecological relations globally and can consider multiple dimensions of interactions between atmospheric, oceanic, and terrestrial processes. Due to the possibilities to record data all over the world, the increase of resolution and quality in recordings from, e.g., satellite platforms, and international efforts to document the global distribution of biodiversity, increasing availability of heterogeneous data sets via the World Wide Web and computing in the cloud, new opportunities arise. These data may enable us to answer questions that are of fundamental importance for the future of our planet. In short: ecology is one of those sciences, affected in a significant way by the tremendous increase in possibilities to collect and analyse data, and there is significant societal interest in taking advantage of these possibilities.

In the following, we will look at the topic from two perspectives. First, from the perspective of ecological research: Where would it profit from computer science? And second, from the perspective of computer science: where could it support ecological research and gain challenging research questions from such a collaboration? We will start with a rather general discussion, but then narrow each topic down to one rather specific problem.

One example discipline, where the 4th scientific paradigm may revolutionize the epistemic foundations could be ecology: Ecologists have been collecting data all over the world and organizational scales ranging from microscopic processes to global phenomena. For instance, latest developments in metagenomics have opened the possibility to prove the occurrence of species across a wide range of taxonomic hierarchies via "Environmental DNA" [1] – several thousands of samples can be collected within reasonable time frames. Satellite remote sensing data offer temporally continuous and spatially contiguous estimates of the states of land and aquatic ecosystems [2]. Monitoring biologically mediated fluxes of CO2 between land and atmosphere exchanges allow monitoring of ecological processes [3] (http://fluxnet.ornl.gov/). Soundscapes of birds [4] offer new ways to determine species diversity. All these examples show that novel observational methodologies are currently revolutionizing this branch of science. In all cases, the resulting data streams are heterogeneous and often unstructured, even when the same processes are observed by different groups, or over different regions of the world. Nevertheless, model building is heavily supported by the collected data. Furthermore, increasingly sophisticated models are developed, which are parameterized or calibrated with different sources of data [5] and demand very substantial computing power. Most information cannot be extracted from the data without computer support during the analysis, storage, access, distribution, visualization.

Besides typical "big-data" problems caused by volume, velocity, variety and veracity of data, there are more important challenges: providing access to the right data (and in an appropriate structure), to extract the relevant information considering redundancies and knowledge, and to develop computationally efficient ways for data model linkages. Therefore, at least three general topic areas can be identified:

Obtaining and Preserving Data

This includes automatic monitoring schemes, automatic interpretation of e.g. remote sensing or image data, sampling bias analysis and gap-filling, data quality management, synthesis and curation. A particular challenge is the huge heterogeneity of data ranging from sequence data to remote sensing images, and from digitized natural history museum collections to manually collected observation data to audio files capturing acoustic diversity. A second important challenge is the increasing volume of such data evident already for remote sensing data and for sequence and related data, where new techniques and rapidly sinking prices lead to an explosion in data volume.

Pattern-recognition in highly dimensional and geo-tagged data sets

The field involves developing sound and efficient algorithms able to capture structure and feature relations in empirical data, and mostly involve finding groups (clustering), anomalies (detection), automatic categorization and prediction (classification/regression), and learning proper representation spaces (visualization) of generally unstructured, heterogeneous, multimodal data streams where quantifying uncertainty is mandatory.

Model development and Model-Data-Confrontation (see e.g. [6])

This includes dealing with sampling bias and scale issues, methods for fitting model to data, scaling and parallelization for cluster or cloud computing.

Some areas of computer science that can contribute to these topic areas and derive research questions from them are:

Data and Model Management

Data Management is certainly the part of computer science that has been used in ecology the longest and is one of the major focus areas of Ecoinformatics. Numerous data management platforms and workflow environments suitable for ecological data have been developed focussing on different stages of data management from data collection in the field (supported, e.g., by smartphone applications) to long term preservation of data. As major challenge remains the seamless integration of data management tasks in the usual workflows of the researchers. A key part of this challenge is identifying what data are useful for particular types of analysis and purposes. Capturing the pragmatic relationships between data and their use, including the tasks and methods for which data have been successfully used, remains a relatively unexplored area of research. Additionally, platforms are needed that can deal with the vast heterogeneity of the data and the expected future huge volumes of data. Increasingly, ecological data of high spatial and temporal resolution can be crowdsourced and streamed from sensors of variable quality, and despite the great potential for this data to be used for ecological analysis the heterogeneity of sources creates open research challenges for data management. New challenges arise also from the vast amount and poor quality of sequencing data; requiring new bioinformatics techniques to handle and preserve the data.

Data Integration

The ability to integrate data is vital for ecological research. However, such integration is hampered by a number of factors where the application of modern approaches from computer science will be helpful. Over the last few years, considerable effort went into the development of formal, machine-readable taxonomies and metadata standards; the use of ontologies is relatively widespread. This requires ontology matching and modularisation. Often, integration problems are present at the instance rather than the schema level. Approaches for duplicate detection and data quality assurance are needed here. Provenance and uncertainty management are needed for gaining meaningful results from the integrated data. This area poses a real challenge for computer science since the information that needs to be encoded goes well beyond the rather simplistic e.g. simple probability distributions commonly used today.

Modern techniques from Computer Vision, Pattern Recognition, Data Mining and Machine Learning

Over the last years, computer vision research already tackled problems that are of high relevance for ecological research as well. One example is the analysis of remote sensing data, which forms one of the basis for global analysis of terrestrial processes, for which several modern methods for automatic processing exist, for example, semantic segmentation. Other examples include large scale analysis of the distribution of animals, plants, and (increasingly genetically derived) populations [7], whereby the data often suffers from extremely biased (in space and time) sampling [8] and few data are available for organism groups where it is difficult to identify the species. Several computer-based methods have recently been developed to support ecological research. These include object recognition software for e.g. plants. However, since those objects offer not just very challenging problems but also call for new methods, that lead to the area of fine-grained recognition. Although today's state of the art systems achieve only recognition rates of 70-80%, in some scenarios machine vision systems are already better than the inexperienced user. Together with techniques from machine learning, like active learning (i.e. keeping the human in the loop as in recent activities), and novelty detection, i.e. detecting if a new object or event is observed, preliminary life-long learning systems are currently under development. In such an iterative manner of building recognition systems and improving performance by specific feedback of users, it is expected that performance of automatic analysis of animals or plants from images and videos will reach the threshold that almost fully automatic observation of our environment will be possible. Having such methods will bring researchers from ecology closer to measurement stations equipped with cameras that could record the environment at a level that has not been possible before. Finally, computer vision techniques might support digitalization of existing ecological data sets. Besides computer vision, modern machine learning techniques will play an important role in the future of ecology data analysis as well. For example, analysing huge amount of data by the human can be supported by automatic clustering into relevant groups. Dimensionality reduction methods, like non-linear or kernel PCA offer new potentials in data pre-processing. Detecting the unexpected, i.e. interesting in data streams can be supported by automatic analysis using novelty and anomaly detection methods, and thus can serve as clustering in the sense of reduction of human efforts to the most important parts of data streams. Finally, machine learning techniques in general might help to make the invisible visible by solving regression problems using training data. Such mappings from input data to output might be the basis for future decision based on measurement. Estimation of bio-geo-chemical parameters using advanced retrieval methods currently provide accurate time-resolved estimations, but advances on uncertainty estimation (going beyond point-wise predictions to meaningful confidence intervals) and knowledge discovery capabilities (i.e. ranking input features to understand the underlying bio-physical processes) are still needed.

High-Performance and Cloud Computing (bring computing power to the data)

The growing amount of data and increasingly complex models require new ways of processing. It is no longer feasible – as is done today – to select data from some online source and download it for local processing. Rather than launching the data to the algorithms, the trend is to launch the algorithms to the data. Here, approaches for function shipping and/or parallelisation can be helpful and are successfully applied, e.g., by GBIF for (re-)ingest of data or in the Map of Life project. Ecological information analysis and modeling largely remains restricted in the size and complexity of problems that can be addressed due to lack of research into up-scaling ecological algorithms (e.g. analysis of ecosystem connectivity) from desktop applications to high performance computing. This requires a systematic approach of mapping ecological data structures and algorithms to well-understood techniques of parallel computation and communication that have been identified by the high-performance computing research community. Identification of how environmental simulations and analyses map to compositions of these well-established scientific computing patterns will be a necessary outcome of this research. Another challenge is model design to best meet recent advances in computer science. This includes, e.g., re-designing models to run on energy-efficient graphics processing units (GPUs). Running models on GPUs instead of conventional CPUs can decrease electricity costs very substantially.

In order to provide a more detailed understanding of some of the problems involved, let us have a look at three concrete examples that highlight different problem areas and different possible links between computer science and ecology.

Example 1: Biodiversity Weather Stations/Automated Long-Term Monitoring

Traditionally, data in ecological research have been collected manually on a rather small scale. For instance, the traditional approach to analysing species richness in a tropical rainforest is to select a plot of manageable size and send scientists (typically PhD students) there, to map the species that occur on this plot. This approach has several drawbacks: First, it is extremely expensive. Second, since neither money nor personnel are unlimited resources, it scales poorly. Third, the quality of the result depends a lot on the expertise of the scientists in the field. The acknowledgements of a recent paper on tree flora in the Amazonian that aims at developing a large scale model and uses data from around 2000 plots, e.g., states “This paper is the result of the work of hundreds of different scientists and research institutions in the Amazon over the past 80 years”, Basically the same drawbacks exist for other types of data collection in ecological research. For instance, in the Biodiversity Exploratories, insect populations on research plots are determined by installing window traps in the field which collect insects. The species are then determined by manual analysis by large numbers of student helpers analysing every caught individual.

In the future, such monitoring schemes could be automated. Technologies like DNA-barcoding of environmental samples, visual and acoustic identification of animals, identification of plants via emitted chemicals are currently being combined to build an Automated Multisensor Station for Monitoring of Species Diversity (AMMOD). The AMMOD requires a combination of image and sound recognition, machine-readable reference libraries for genetic and biochemical markers, images and sounds, the storage and sorting of a large amounts of data and finally, when several stations are combined, modeling of species distribution in landscapes.

Example 2: Global Change Ecology

Key challenges for Ecology in our Global Change era are i.) to understand and predict the geographical distributions and abundances of species and populations and ii.) to improve our understanding of the role of biodiversity for the functioning of ecosystems [11] and their supply of services to the human society under Global Change. Addressing these challenges implies dealing with spatially biased data, e.g. for the occurrence of species, and integrating various data types on where species or populations occur, which functional traits they have, the environment in which they live (e.g. climate, soil types, land cover) and ecosystem processes, such as biomass productivity and carbon cycling [12]. Thus, it is necessary to integrate multiple types of data from the biological and geosciences, ranging from genetic data characterising populations or species to satellite-derived estimates of land cover change [13]. Thereby, the genetic and satellite data, in particular, have reached levels of complexity and sizes, which are sometimes beyond the capacities of normal desktop computers. Instead, massive RAM or parallel cluster computing are increasingly necessary to handle the data, even for relatively simple analyses. For more complex model-data fusion techniques, such as hierarchical Bayesian modeling, computational capacities are still highly limiting ecological research.

Example 3: Modelling ecosystem and Earth system processes

Modelling now also plays a crucial role for ecosystem science from the local to global scale. More and more ecological processes are currently integrated into so-called Earth System models, which integrate climate models with biosphere models [14, 15]. Yet, there is a large uncertainty in future model predictions for these dynamic systems [16]. One challenge now is to provide observation-based constraints which can confine future model behaviour. We need to understand better which patterns of the observations provide robust constraints for models. Hence, we need to move away from simple model-data comparisons, to pattern-oriented model evaluation, calibration and interpretation in a system-oriented way [17]. Examples of this include approximate Bayesian computation [18] and the concept of emerging constraints [19]. As a variety of data types, ranging from leaf-level measurement of photosynthesis to satellite-derived estimates of forest biomass, can be used to parameterize and constrain ecosystem models, such models might in the future rather serve as process-based linkages between multiple data types, instead of just being parameterized and tested with individual data sets at a time. At lot of analogies between video data and dynamic Earth System data have been identified and ideas generated of how applying methods of one domain in the other.

In summary, we strongly believe that a closer interaction between ecologists and computer scientists is needed to tackle the challenges in Ecology and that both disciplines will profit from such interaction: Ecologists will be able to solve problems currently beyond their reach. Computer Scientists will be exposed to a challenging set of real-world problems requiring the development of new methods and approaches.

References

1. P. Taberlet, E. Coissac, M. Hajibabaei, L. H. Rieseberg, "Environmental DNA", Molecular Ecology, 21(8), 1789–1793, 2012.
2. M.-N. Tuanmu, W. Jetz, "A global 1-km consensus land-cover product for biodiversity and ecosystem modelling", Global Ecology and Biogeography 23:1031–1045, 2014.
3. D. Baldocchi, "Measuring fluxes of trace gases and energy between ecosystems and the atmosphere – the state and future of the eddy covariance method", Global change biology, 20(12):3600–3609, 2014.
4. E. P. Kasten, S. H. Gage, J. Fox, W. Joo, "The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology", Ecological Informatics, 12, 50–67, 2012.
5. F. Hartig, J. Dyke, T. Hickler, S. I. Higgins, R. B. O'Hara, S. Scheiter, A. Huth, "Connecting dynamic vegetation models to data – an inverse perspective", Journal of Biogeography, 39(12), 2240–2252, 2012.
6. M. C. Rillig, W. Kiessling, T. Borsch, A. Gessler, A. D. Greenwood, H. Hofer, ..., F. Jeltsch, "Biodiversity research: data without theory – theory without data", Frontiers in Ecology and Evolution, 3, 20, 2015.
7. M. Balint, M.S. Domisch, C. H. M. Engelhardt, P. Haase, S. Lehrian, J. Sauer, K. Theissinger, S. U. Pauls, and C. Nowak, "Cryptic biodiversity loss linked to global climate change", Nature Clim. Change, 1:313–318, 2011.
8. C. Meyer, H. Kreft, R. Guralnick, W. Jetz, "Global priorities for an effective information basis of biodiversity distributions", Nature Communicatins, 6:8221, 2015.
9. EU COST Action: "Mapping and the citizen sensor", http://www.citizensensor-cost.eu/
10. H. Ter Steege, N. C. Pitman, D. Sabatier, C. Baraloto, R. P. Salomão, J. E. Guevara, ..., P. V. Fine, "Hyperdominance in the Amazonian tree flora", Science, 342(6156), 1243092, 2013.
11. F. T. Maestre, J. L. Quero, N. J. Gotelli et al., "Plant Species Richness and Ecosystem Multifunctionality in Global Drylands", Science, 335:214–218, 2012.
12. H. M. Pereira, S. Ferrier, M. Walters, G. Geller, R. Jongman, R. Scholes, M. W. Bruford, N. Brummitt, S. Butchart, A. Cardoso, "Essential biodiversity variables", Science, 339:277–278, 2013.
13. M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, J. R. G. Townshend, "High-Resolution Global Maps of 21st-Century Forest Cover Change", Science, 342:850–853, 2013.
14. G. B. Bonan, S. Levis, S. Sitch, M. Vertenstein, K. W. Oleson, "A dynamic global vegetation model for use with climate models: concepts and description of simulated vegetation dynamics", Global Change Biology, 9(11):1543–1566, 2003.

15 P. M. Cox, R. A. Betts, C. D. Jones, S. A. Spall, I. J. Totterdell, "Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model", Nature, 408: 184–187, 2000.
16 M. Heimann, M. Reichstein, "Terrestrial ecosystem carbon dynamics and climate feedbacks", Nature, 451: 289–292, 2008.
17 M. Reichstein, C. Beer, "Soil respiration across scales: The importance of a model-data integration framework for data interpretation", Journal of Plant Nutrition and Soil Science, 171(3):344–354, 2008.
18 J. A. Vrugt, M. Sadegh, "Toward diagnostic model calibratn and evaluation: Approximate Bayesian computation", Water Resources Research 49.7 (2013): 4335–4345.
19 P. M. Cox, D. Pearson, B. B. Booth, P. Friedlingstein, C. Huntingford, C. D. Jones, C. M. Luke, "Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability", Nature 494, no. 7437 (2013): 341–344.

4 Overview of Talks

4.1 Life-long Learning with Applications in Monitoring Biodiversity

Joachim Denzler (Universität Jena, DE)

Joint work of Joachim Denzler, Erik Rodner, Alexander Freytag, Christoph Käding, Marcel Simon, Clemens-Alexander Brust

Most of today's impressive results in computer vision and machine learning arise from two major changes during the past 20 years: Firstly, the increased performance of hardware together with the advent of powerful graphical processing units (GPU) applied in scientific computing beyond pure displaying. Secondly, the huge amount of, in part, annotated image data provided by today's generation of Facebook and Twitter users and available easily over databases (e.g., Flickr) and/or search engines. Consequently, tasks like face recognition and identification can be solved using powerful methods, like Convolutional Neural Networks [13], and millions of face images for training.

For visual monitoring of biodiversity, for example, to keep track of species distribution of certain mammals, no such databases or collection of annotated or even weakly annotated images exists at a size such that systems can be directly trained. This first challenge, the collection of training data bases for computer vision algorithms, links directly to the citizen science activities. We need to motivate people to also share their annotated images of animals, insects, and other species, or at least to help collecting such databases.

Although training data is now one limiting factor for visual monitoring at a certain level of quality, there are several other and equally important challenges from the computer vision and machine learning perspective:

1. Number of species to be distinguished: although current computer vision systems can differentiate between up to 10.000 different categories (see ImageNet [2]), this number is far from being sufficient for the number of species to be expected in Germany. In addition, the classification of such many different objects has been demonstrated at the category level only, i.e. to differentiate dogs from cats, but not certain races of dogs and cats.
2. Generic classifiers: although certain systems already exist for analyzing images of moths, chimpanzees, or other specific class of objects [12, 4], those systems have been carefully developed using handcrafted and optimized features and individual domain knowledge. At

present, it seems not possible that such specialized system can be individually developed for all the different classes of animals and insects to be monitored. Thus, there is the need for generic classifiers that learn their feature representation from data, at best in an unsupervised manner [1, 11].

3. Fine-grained recognition: As mentioned earlier, most computer vision system for classification of objects in an image, are already powerful if it comes to distinction of categories, like cups, cars, dogs, etc. Within category classification, i.e. the distinction between a Great Spotted Woodpecker and a Middle Spotted Woodpecker, it is a much more challenging problem, and currently the focus in fine-grained recognition. For certain categories, like birds, cats, and dogs, solutions are already available [3]. However, there is still a generic method missing that identifies the relevant, visual parts of objects that allow reliable classification within a category of visually similar species.
4. Detection of the unexpected: Today's machine learning system work under the closed-world assumption, i.e. they will map any input image to one of the known classes. Species not known to the system will not be correctly classified, but even worse might be wrongly assigned to a known class. Since the unexpected is often the driver of progress in science, such wrong assignments might prevent some insight in the monitored ecosystem. Thus, methods for novelty and anomaly detection is another big challenge to not miss the probably important insight from unexpected observations [5, 10].
5. Keeping the human in the loop: Today, it cannot be expected that automatic monitoring systems will work error-free from scratch. The challenge arising from difficult and changing recoding conditions in the wild, hiding and only partially visible animals will result in erroneous assignment or even misses of objects visible in the image for the human. Thus, acceptance of such systems in the monitoring community will heavily depend on reliability of the automatically generated statistics and properties of the observed species. Consequently, one additional challenge is to provide a feedback mechanism from the machine to the human, to report about uncertain or undetermined results. However, the feedback from the human to the machine is equally important by correcting results or adding additional information for refinement and optimization of the automatic system [7].

In summary, we believe that automatic visual monitoring should be framed in a life-long learning cycle that has been recently applied to monitor mammals in Portugal [6]. The key ingredients are initial, generic classifier, for example, powerful CNN architectures [13], active learning to reduce costly annotation effort by experts [8, 5], fine-grained recognition to differentiate between visually very similar species [3], and efficient incremental update of the classifier's model over time [9]. For most of these challenges, initial solutions exist. Building first visual monitoring systems, possibly for a restricted area or set of species, will definitely help to improve all parts over time, if biodiversity and computer vision researchers are working closely together.

References

1 A. Freytag, E. Rodner, M. Simon, A. Loos, H. Kühl, J. Denzler, "Chimpanzee Faces in the Wild: Log-Euclidean CNNs for Predicting Identities and Attributes of Primates", German Conference on Pattern Recognition (GCPR), 51–63, 2016.

2 J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database", International Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

3 M. Simon, E. Rodner, "Neural Activation Constellations: Unsupervised Part Model Discovery with Convolutional Networks", International Conference on Computer Vision (ICCV), 2016.
4 J. Balfer, F. Schöler, V. Steinhage, "Semantic Skeletonization for Structural Plant Analysis", 7th Intern. Conf. on Functional-Structural Plant Models (FSPM 2013), Saariselkä, Finland, June 9-14, 2013.
5 Ch. Käding, A. Freytag, E. Rodner, P. Bodesheim, J. Denzler, "Active Learning and Discovery of Object Categories in the Presence of Unnameable Instances", IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 4343–4352, 2015.
6 Ch. Käding, E. Rodner, A. Freytag, J. Denzler, "Watch, Ask, Learn, and Improve: A Lifelong Learning Cycle for Visual Recognition", European Symposium on Artificial Neural Networks (ESANN), 2016.
7 Ch. Käding, A. Freytag, E. Rodner, A. Perino, J. Denzler, "Large-scale Active Learning with Approximated Expected Model Output Changes", German Conference on Pattern Recognition (GCPR), 179–191, 2016.
8 A. Freytag, E. Rodner, J. Denzler, "Selecting Influential Examples: Active Learning with Expected Model Output Changes", European Conference on Computer Vision (ECCV), 562–577, 2014.
9 E. Rodner, A. Freytag, P. Bodesheim, B. Fröhlich, J. Denzler, "Large-Scale Gaussian Process Inference with Generalized Histogram Intersection Kernels for Visual Recognition Tasks", International Journal of Computer Vision (IJCV), 253–280, 2017.
10 P. Bodesheim, A. Freytag, E. Rodner, M. Kemmler, J. Denzler, "Kernel Null Space Methods for Novelty Detection", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3374–3381, 2013.
11 E. Rodner, M. Simon, G. Brehm, S. Pietsch, J. W. Wägele, J. Denzler, "Fine-grained Recognition Datasets for Biodiversity Analysis", CVPR Workshop on Fine-grained Visual Classification (CVPR-WS), 2015.
12 A. Loos, A. Ernst, "An automated chimpanzee identification system using face detection and recognition", EURASIP journal on image and video processing, 2013.
13 Y. LeCun, J. Bengio, G. Hinton, "Deep Learning", Nature 521, 436–444. doi:10.1038/nature14, 2015.

4.2 Systematic Evaluation of Land Surface Models Using the International Land Model Benchmarking (ILAMB) Package

Forrest Hoffman (Oak Ridge National Laboratory, US), Nathaniel O. Collier, Gretchen Keppel-Aleks, Charles D. Koven, David M. Lawrence, Mingquan Mu, James T. Randerson, William W. Riley

Main reference F. M. Hoffman et al., "2016 International Land Model Benchmarking (ILAMB) Workshop Report", Technical Report DOE/SC-0186, 2017.
URL http://dx.doi.org/10.2172/1330803
URL http://dx.doi.org/10.18139/ILAMB.v002.00/1251621

As Earth system models (ESMs) become increasingly complex, there is a growing need for comprehensive and multi-faceted evaluation of model predictions. To advance understanding of biogeochemical processes and their interactions with hydrology and climate under conditions of increasing atmospheric carbon dioxide, new methods are needed that use observations to constrain model predictions, inform model development, and identify needed measurements and field experiments. Improved process parameterizations are needed to constrain energy and water predictions in land surface models and better representations of biogeochemistry – climate feedbacks and ecosystem processes in ESMs are essential for reducing uncertainties associated with projections of climate change during the remainder of the 21st century. The International Land Model Benchmarking (ILAMB) project seeks to 1) develop internationally accepted benchmarks for land model performance, 2) promote use of benchmarks for model intercomparison projects, 3) strengthen linkages between experimental, remote sensing, and modeling communities, and 4) support the design and development of an open source benchmarking software system. Leveraging work on past model evaluation studies, we have developed two generations of such benchmarking software packages that assess model fidelity on 24 variables in four categories from about 45 data sets; produce graphical global-, regional-, and site-level diagnostics; and provide a hierarchical scoring system. The ILAMBv2 package, publicly released in May 2016, has become an integral part of model verification workflow for rapid model development and calibration cycles for the U.S. Department of Energy's Accelerated Climate Modeling for Energy (ACME) model and the Community Earth System Model (CESM). We will present results from model analysis using the ILAMB packages, discuss techniques for routine model evaluation, propose coordinated evaluation of the Sixth Phase of the Coupled Model Intercomparison Project (CMIP6) output, and describe new metrics that integrate across carbon, surface energy, hydrology, and land use disciplines.

4.3 BExIS 2 – An open source data management platform for collaborative projects in Biodiversity Research and beyond

Birgitta König-Ries (Universität Jena, DE)

URL http://bexis2.uni-jena.de

In many collaborative projects, there is a strong need for data preservation and sharing. BExIS 2 is an open source data management platform that meets these needs and supports data management throughout the entire data lifecycle. It is a modular platform that can easily be adapted to the specific needs of particular projects with respect to, e.g., access rights, data structure, or metadata schema used. Further information including an online demo and a download link can be found on the BExIS 2 website: http://bexis2.uni-jena.de.

4.4 Real time monitoring of vegetation phenology with the PhenoCam network

Andrew Richardson (Harvard University – Cambridge, US)

Phenology – the seasonal rhythms of plants and animals – has been shown to be a robust integrator of the effects of year-to-year climate variability and longer-term climate change on natural systems. At the level of ecosystems, phenology is important because it influences productivity, carbon sequestration, nutrient cycling, and feedbacks to the atmosphere and climate system.

There is a demonstrated need to better document biological responses to a changing world, and improved phenological monitoring will contribute to achieving this goal. In this talk, I will describe a collaborative research network called "PhenoCam" (http://phenocam.sr.unh.edu/). PhenoCam uses networked digital cameras – webcams – for phenological monitoring in a range of ecosystems (almost 400 sites, and 750+ site-years of archived data) across the North American continent. Images are captured every 30 minutes, uploaded to the PhenoCam server for display in real-time, and processed to yield quantitative measures of vegetation "greenness." I will conclude by talking about some of the challenges we face with managing this ever-expanding image archive.

4.5 New technologies for biodiversity monitoring

Johann Wolfgang Wägele (ZFMK – Bonn, DE)

Biodiversity is one of the most valuable resources of our planet. With possibly more than 10 million living species and most of these still unknown to science, the biosphere of our planet guarantees future generations a wealth of hitherto untapped genetic resources, which are relevant for food production, medicine, bioenergy production, and life-supporting ecosystem functions. In contrast to global warming, a steady loss of biodiversity is irreversible and leads to an impoverished world that will not recover its original richness within the next 5 million years. Already more than 20 years ago the large-scale destruction of habitats and losses of biodiversity alarmed researchers and policy makers. Until today, the biodiversity crisis is accelerating and a trend reversal is not achievable with political treaties and resolutions solely. One important reason is the lack of reliable high resolution, large scale data. Such data are needed as a basis for informed decisions, to analyze causes of local extinctions, to prove that trends are really happening, to model scenarios that explain ongoing changes and that can predict future processes, and to define actions based on scientific information. In analogy to climate scientists, who were able to raise awareness for ongoing climate changes at a global scale, biologists need data to advice policy makers, to convince stakeholders and the general public. The most significant impediment for large-scale and fine-grained biodiversity monitoring is the taxonomic one. Even when sampling campaigns are well planned and executed, the samples have little value if the majority of species cannot be identified. This difficulty is mainly due to the lack of time to sort and identify all species found, combined with the fact that taxonomist are scarce and largely specialized for selected taxa, which again makes the majority of identifications very time-consuming and not doable by untrained ecologists. Another problem is that monitoring schemes usually are not comparable, and programs do not run long enough to document trends. Climate monitoring using satellite images and automatized weather stations has been organized at a large scale everywhere on earth. In contrast, large-scale and long-term monitoring of biodiversity does not exist, among others, because the required technology has not been developed. It is therefore crucial to adapt existing technologies for the development of automatized biodiversity motoring. We need "weather stations for species diversity". It is possible to construct an automatized multisensor station for monitoring of species diversity (an AMMOD) using already available technology: bioacoustics sensors, automated image analyses, DNA-barcoding, analyses of volatile organic compounds (VOCs). Thus it is possible to detect mammals and birds (mainly via images and sounds), insects (mainly with DNA barcoding), plants (via barcoding of pollen and VOCs), and soil microorganisms (via emitted VOCs). AMMODs allow for a continuous detection of a large number of species, Main challenges are in the field of computer science: pattern recognition and comparison of environmental signals with reference databases has to be improved to increase resolution.

5 Working groups

5.1 Biodiversity Weather Stations

Tilo Burghardt (University of Bristol, GB), Yun-Heh Jessica Chen-Burger (Heriot-Watt University – Edinburgh, GB), Joachim Denzler (Universität Jena, DE), Birgitta König-Ries (Universität Jena, DE), Miguel Mahecha (MPI für Biogeochemistry – Jena, DE), Shawn Newsam (University of California – Merced, US), Natalia Petrovskaya (University of Birmingham, GB), and Johann Wolfgang Wägele (ZFMK – Bonn, DE)

The working group focused on the question how an increasing need for data availability on global biodiversity information can be met by introduction of automated field stations for in-habitat sampling. The discussion was underpinned by previous conceptual work on a concept laid down in proposals for BioM-D (Deutsches Zentrum fuer Biodiversitätsmonitoring)/AMMOD (Automatic Multi-sensory station for Monitoring Of species Diversity).

Motivation

The group emphasised that biodiversity is one of the most valuable resources of the planet; and that changes due to species extinction are irreversible. Monitoring biodiversity must therefore be a key component and precursor for taking informed decisions about ecosystem management and conservation. Currently, major obstacles prevent large-scale monitoring of biodiversity at species level: 1) the difficulty of taxonomic identification, 2) the difficulty of spatial-temporal coverage, 3) the difficulty of meaningful spatio-temporal reference, and 4) the workload problem: automatic workflows are in their infancy.

Concept

Faced with these impediments, the group supported the concept that, to be able to observe global change of our biosphere, we need an infrastructure comparable to that used by climate researchers; that is 'weather stations for species diversity'", which operate in a similar fashion to traditional weather stations sampling the breadth of species presence at a particular sampling location over time. In fact, biologists have started to adopt various technologies to enable such measurements – bringing these technologies together for an automatic multi-sensory station for monitoring of species diversity establishes a clear, interdisciplinary development goal. The group reiterated previously identified candidate modalities for automated monitoring; these include DNA barcoding, bio-acoustic monitoring, computer vision-based surveillance, and the analyses of 'smell-scapes'.

Discussions

Driven by the difficulties experienced in raising appropriate funding to progress this agenda at scale, the group discussed proof-of-concept options for the introduction of a few prototypical stations that could demonstrate the value and practical operation of the concept first hand. Key conclusions here included the identification of existing tower infrastructures for the commissioning of systems, the focus on well established sites for best cross-referencing of data, and the limitation of developments for a few species most relevant for showcasing the capabilities of a prototype. We also discussed the technologies underpinning these stations and practical ways of utilising the expertise of scientists and research groups to best conduct development work towards the establishment of prototypes.

Conclusions

The group concluded to work on a detailed positioning paper that may include authorship of the wider community over the following months, and continued efforts towards funding of the concept as next steps. The working group made clear that the technological foundations for an AMMOD concept are widely available today, and that a strong effort is needed to turn this foundation into a practical, working infrastructure to support the gathering of biodiversity information at scale.

5.2 Blending machine learning methods and process-based approaches in dynamic ecological models

Florian Hartig (Universität Regensburg, DE), Martin Bücker (Universität Jena, DE), Gustau Camps-Valls (University of Valencia, ES), Forrest Hoffman (Oak Ridge National Laboratory, US), Kazuhito Ichii (JAMSTEC – Yokohama, JP), Martin Jung (MPI für Biogeochemistry – Jena, DE), Bertram Ludäscher (University of Illinois at Urbana-Champaign, US), Markus Reichstein (MPI für Biogeochemistry – Jena, DE), and Jakob Zscheischler (ETH Zürich, CH)

In many areas of ecology and earth-system sciences, process-based computer simulations are central for understanding the dynamic response of ecological systems to external forcings. An example are dynamic vegetation models, which describe the response of vegetation ecosystems, typically represented by soil water, nutrient and plant state variables, to disturbances and climatic forcings (Zaehle & Friend, 2010; Forkel et al. 2016). The processes governing these dynamics are often complex, and can only partially be observed. It has therefore become common to statistically calibrate model parameters to field observations, for example to vegetation inventories, measured gas exchange, or remote-sensing data (e.g. Hartig et al., 2012).

The issue with this approach is that statistical methods, while accounting for the fact that parameter values are uncertain and some stochastic error is present, are contingent on the assumption that the underlying data-generating model is correct. In other words, statistical conclusions are generally only correct if the fitted model is approximately correct. While this assumption is of lesser concern in simple regression problems, it becomes a major concern in complex, dynamic models, where errors may propagate through nonlinear processes into other model compartments or times, which can create a range of problems for the correct estimation of model parameters, their uncertainty, and associated forecasts.

An obvious solution is making the structure of process-based models more flexible. This could mean, for example, that process-descriptions, which are typically specified by relatively rigid formulae, are replaced by a flexible statistical approach. An early attempt at implementing such an approach is Wood et al. (2001), who replaced fixed formulae by flexible generalized additive models (see also Nisbet et al., 2004). Another possibility is to make the model structure itself flexible, by adding or removing state variables, or their connections (e.g. Babtie et al., 2014).

In this working group, we discussed those and other technical approaches to tackle the problem of creating flexible models that blend machine learning and process-based models. In particular, we considered the problem of a complex dynamic system, where very little prior information about a particular subprocess is available. The challenge is thus to train a flexible statistical algorithm to learn the dynamical response of the subprocess from observing the system as a whole, while at the same time keeping the problem computationally tractable. A possible solution identified by the group was the use of automatic differentiation methods (Griewank and Walther, 2008), which seemed promising for creating a computationally efficient blend of process-models with machine-learning methods.

Acknowledgements: the working group would like to acknowledge useful suggestions from Shawn Newsam and Andrew Richardson.

References

1. A. C. Babtie, P. Kirk, M. P. H. Stumpf (2014). *Topological sensitivity analysis for systems biology.* Proceedings of the National Academy of Sciences 2014 111 (52), 18507–18512
2. M. Forkel, N. Carvalhais, C. Rödenbeck, R. Keeling, M. Heimann, K. Thonicke, S. Zaehle, M. Reichstein (2016). *Enhanced seasonal CO2 exchange caused by amplified plant productivity in northern ecosystems.* Science 351, 696–699.
3. A. Griewank, A. Walther (2008). *Evaluating Derivatives.* Society for Industrial and Applied Mathematics.
4. F. Hartig, J. Dyke, T. Hickler, S. I. Higgins, R. B. O'Hara, S. Scheiter, A. Huth (2012). *Connecting dynamic vegetation models to data – an inverse perspective.* Journal of Biogeography 39 (12), 2240–2252.
5. R. M. Nisbet, E. B. Muller, K. Lika, S. A. L. M. Kooijman (2000). *From molecules to ecosystems through dynamic energy budget models.* Journal of Animal Ecology 69 (6), 913–926.
6. S. N. Wood (2001). *Partially specified ecological models.* Ecological Monographs 71(1), 1–25.
7. S. Zaehle, A. D. Friend (2010). *Carbon and nitrogen cycle dynamics in the o-cn land surface model: 1. model description, site-scale evaluation, and sensitivity to parameter estimates.* Global Biogeochemical Cycles 24, 1–13.

5.3 Improving data discovery and integration for global ecological analyses

Ivaylo Kostadinov (Jacobs University Bremen, DE), Martin Bücker (Universität Jena, DE), Matthew Evans (University of Hong Kong, HK), Thomas Hickler (Senckenberg Research Centre, DE), Donald Hobern (GBIF – Copenhagen, DK), Birgitta König-Ries (Universität Jena & iDiv, DE), Bertram Ludäscher (University of Illinois at Urbana-Champaign, US), Frank Pennekamp (Universität Zürich, CH), Brody Sandel (Santa Clara University, US), and Bernhard Seeger (Universität Marburg, DE)

This working group collected key challenges in contemporary ecological research, which are based on different aspects of data, including discoverability, standardization, integration, size, and metadata description. Current and future approaches in addressing these challenges were discussed in a dialogue between scientists covering the whole range between ecology and informatics. In the later part of the seminar the group focused on the (semi-)automated distribution and harvesting of data and metadata for ecological analyses.

Motivation

Understanding ecological systems, for example detecting biodiversity change and pinpointing its main drivers, is a major challenge which nowadays requires close cooperation between ecology and computer science to tackle. Studying global patterns requires the acquisition, management and integration of large, heterogeneous datasets, whose complexity increases rapidly. Heterogeneity is among the key challenges in working with biodiversity data. Differences in acquisition protocols, study areas and strategies in dealing with gaps in the observations are only some of the problems in integrating data from different sources, disciplines and scale.

Discussion

Ecological communities can be described by four main data types (abundance, distribution, traits, genetic) and their environment is described by direct measurements and remote-sensing approaches. One of the key challenges is trying to bring together different layers of biodiversity change and some of the drivers of that change. For example, traits can constrain species distribution and are therefore one key factor to explore. However, not many dedicated databases for trait data exist and linking to other data (sources) is often difficult. Optimally, there would be a single point for searching and doing at least basic analysis. One possibility would be including traits (e.g. habitat preference) in distribution plots or mapping traits instead of species distributions. This could be attempted in the near future with the GFBio Visualization, Analysis and Transformation (VAT) system (https://vat.gfbio.org) which already offers different plots of publicly available data resources like the Global Biodiversity Information Facility (GBIF, www.gbif.org). Including habitat, temporal, climate continuity in analyses was also identified as a desirable future outcome. This could also be achieved with VAT, by developing a spatio-temporal database in cooperation with scientists, for example to determine apparent correlations between species co-occurrence maps, based on GBIF data. GBIF also strives to improve the data it offers for re-use, for example by including the taxonomic target and completeness as part of the metadata of the sampling event, so that presence/absence can be inferred more correctly.

Data discovery and integration. A common example of the difficulties researchers face when searching and integrating ecological datasets is that even basic metadata like units of measurement is often not available or impossible to compare. Therefore, dedicated tools to systematically extract candidate metadata (e.g. locations, times) from journal texts would be very helpful. A solution, currently employed by GBIF and others, is a wizard-style interface for authors to input basic metadata. However, minimal requirements almost always differ in scope and detail, as shown by some of the most widely used frameworks: Ecological Metadata Language (EML,https://knb.ecoinformatics.org/#tools/eml), Dublin Core (http://dublincore.org/), Darwin Core (DwC, http://rs.tdwg.org/dwc/). In general, one could distinguish between rich and complex data (e.g. most XML-Schema based data) and simple, flat records. One idea to obtain a rich information network is to increase the semantic load of lightweight DwC properties. Some of the main emerging themes to work on the next years were identified to be (1) the development of an integrated data ecosystem, where global players like GBIF provide highly linked data, (2) the automatic extraction of descriptions for curation, integration and (3) increasing the incentives for data producers to deliver high-quality, standardized metadata, for example by introducing or supporting alternative or additional measures of scientific contribution (e.g. micro-crediting system for data publication).

Metadata enrichment. Metadata is essential for the correct integration of data from different resources. However, it is often not readily available or suffers from lack of standardization. The group outlined a multifaceted approach for addressing this challenge. The main roles would be the original data producer, the later data consumer and any software components for automated metadata extraction. The usual workflow was identified to be: the data producer publishes the data together with the associated journal article (some of the metadata might reside in the article or its appendices). Therefore, the possible sources to extract additional metadata from are (1) author-provided metadata, (2) the data itself, (3) the paper. Optimally, part of the metadata will provide (machine readable) links and to further literature, related projects, funding sources, and involved institutions. A combination of different methodologies was deemed as the best way to cover the different aspects of metadata enrichment. One possibility is to use machine learning techniques for text mining and deducing the research domain context, improving user interfaces for user annotation, curation, and validation, all the while applying established terminologies to standardize the outcome. Finally, a feedback-loop to feed back manually annotated and approved content back to the automation steps would improve their performance.

Conclusion and Outlook

The major challenges for improving descriptive metadata of datasets, and consequently their discoverability and interoperability, are (1) providing the right tools and the right incentives for the data producers to provide the metadata in a standardized way, (2) determining the minimal set of parameters, required for interoperability and (3) providing the tools to harvest the required metadata from available resources like the data itself, the corresponding journal article or even program code automatically.

Several informatics techniques like machine learning offer promising solutions for increasing the automation of metadata extraction. In order for them to be applied meaningfully, priorities for improved information capture must be identified. This includes determining

what essential biodiversity variables ought to be captured. Recording of re-usable workflows from user operations, as already employed by VAT, can deliver some insight to the way scientists (re-)use biodiversity data.
Further improving and standardizing the interfaces for data exchange between data repositories, taking into account emerging serialization formats and access means (e.g. protocols), is important. Large, integrated data resources like GBIF will continue to play a key role in paving the road to Linked Open Data for ecology and biodiversity research.
Raising community awareness to the problems at hand is important. This includes clarifying the added-value of high quality metadata and supporting an appropriate credit system for data publication.
As a continuation of the efforts initiated during the seminar, the working group will organize a Hackathon with the following preliminary topics in mind:

- Rapid prototyping tools for automatically extracting metadata
- Integration of different metadata schema
- Exploring different techniques, e.g. deep learning and traditional statistics, and data sources, e.g. Catalogue of Life (www.catalogueoflife.org) for improving data discovery and linkage.

6 Panel discussions

6.1 Reproducibility and teaching needs – a dialogue between ecology and computer science

Frank Pennekamp (Universität Zürich, CH), Martin Bücker (Universität Jena, DE), Tilo Burghardt (University of Bristol, GB), Gustau Camps-Valls (University of Valencia, ES), Yun-Heh Jessica Chen-Burger (Heriot-Watt University – Edinburgh, GB), Joachim Denzler (Universität Jena, DE), Matthew Evans (University of Hong Kong, HK), Florian Hartig (Universität Regensburg, DE), Thomas Hickler (Senckenberg Research Centre, DE), Donald Hobern (GBIF – Copenhagen, DK), Forrest Hoffman (Oak Ridge National Laboratory, US), Kazuhito Ichii (JAMSTEC – Yokohama, JP), Martin Jung (MPI für Biogeochemistry – Jena, DE), Birgitta König-Ries (Universität Jena, DE), Ivaylo Kostadinov (Jacobs University Bremen, DE), Bertram Ludäscher (University of Illinois at Urbana-Champaign, US), Miguel Mahecha (MPI für Biogeochemistry – Jena, DE), Laetitia Navarro (iDiv – Leipzig, DE), Shawn Newsam (University of California – Merced, US), Natalia Petrovskaya (University of Birmingham, GB), Markus Reichstein (MPI für Biogeochemistry – Jena, DE), Andrew Richardson (Harvard University – Cambridge, US), Ribana Roscher (Universität Bonn, DE), Brody Sandel (Santa Clara University, US), Bernhard Seeger (Universität Marburg, DE), Johann Wolfgang Wägele (ZFMK – Bonn, DE), and Jakob Zscheischler (ETH Zürich, CH)

Abstract

This document summarizes the plenary discussion about reproducibility and teaching needs during the seminar. The discussion started with perspectives from ecologists and computer

scientists about the state of the art and challenges of reproducibility in their fields. A key question discussed was the time-scale of reproducibility, which may pose large challenges for computational researchers. On the practical side, data and code archiving practices were discussed, and efforts to provide incentives for reproducible research highlighted. The second part of the discussion covered whether there is a need for joint training efforts by ecologists and computer scientists to generate the next generation of eco-informaticians trained for the challenges of a largely data-driven science. We concluded the plenary discussion with agreement that both disciplines would benefit from a better dialogue.

Reproducibility in ecology and computer science

Ecology is increasingly becoming a data-driven science and hence ecologists need to work with large, complex datasets for which they often lack the appropriate training [2]. This can lead to issues with reproducibility, which is not unique to the field of ecology, but for science in general [4].

The discussion was started with perspectives from ecologists and computer scientists. The available ecologists believed that a large fraction of papers currently published is not fully reproducible, the attitude toward reproducibility is changing within the field. This is partly due to tools like the statistical computing environment R [5] which is increasingly used for analyses in ecology, biology and the life sciences. It is also due to the rise of literate programming tools in R such as knitr and Rmarkdown [3], which allow to intersperse code and text and are increasingly used.

The change is also fostered by changing journal policies that increasingly require data archiving [9] and also the provisioning of computer code for modeling/simulation studies and methods development [1]. Nowadays most of the major journals in ecology and evolution require data archiving for publication, and guidelines and best practices how to make data available can be found [10]. Nevertheless, current practice is still lacking behind [6]. Whereas many of the modern tools for reproducible research were developed by computer scientists (e.g. version control, unit tests, code review), the available computer scientists in the audience expressed doubts whether the practice of reproducibility is actually better developed within their field [4].

A major discussion topic was what reproducibility actually implies. Whereas data archiving and providing scripts may guarantee the correctness and validity of the results at the time of publication, it is not guaranteed that results will be reproducible in the future (e.g. in ten years time). A major challenge is, for instance, the use of high performance computing in many fields of computer science, including computer vision. Ideally, information about software and hardware architecture should be preserved to reproduce results in the future, but this is often prohibitive. Whereas these issues require careful consideration, they probably only concern a small fraction of ecologists, whose research questions still can be addressed on common desktop or laptop computers most of the time.

Towards more reproducible research

Several guidelines for better reproducible results are available [11, 7]. A first step is better data archiving practices. Several recent publications highlight the value of data archiving and give practical advice how to prepare data for long-term archiving. In addition, journals could require that the scripts used for data analysis are submitted for peer review and check that the output of the scripts corresponds to the results reported in a paper [3]. The Association for Computing Machinery, for example, uses a system in which badges are assigned for research results which can independently be reproduced. In the Life Sciences the ReScience Journal aims to publish independent reproductions of published research. They implemented a fully transparent review process in which the reproduction is first peer-reviewed and then published online, providing researchers with incentives to reproduce other's work.

Another possible avenue for improved reproducibility is the use of work flows that document data input and provenance [8]. These work flows often produce visual representations of data sources and processing steps, which can be understood without knowing the data processing language itself.

Students and researchers should also be exposed to reproducible research early on in their careers. One possible way to do so is a reproducible research journal club. Instead of just reading and discussing research papers, the goal is to reproduce the analysis or model of a given paper. To do so, the students have to access the data from a publicly available source such as Dryad or contact the authors directly. Consequently, the analysis of the paper is performed independently from the authors of the study, just based on the information provided in the paper. Ideally, one does not only reproduce the results of the study, but also learns and understands a method, and learns about the steps a researcher has taken during the analysis. An example of such a reproducible research journal club is run by Owen Petchey at the University of Zurich[1], with successful reproductions publicly available: http://opetchey.github.io/RREEBES/.

Training needs for better computing practices in ecology?

The challenges in managing increasingly large and complex datasets require appropriate training of ecologists. Two non-profit organizations are dedicated to provide training for scientific computing and data management: Software Carpentry[2] is primarily dedicated to train scientists and engineers basic principles to make their scientific computing applications reliable. Data Carpentry[3] on the other hand focuses on providing training in teaching the basic skills to conduct data-driven research such as cleaning, integrating, managing and visualizing large datasets covering the full data life cycle for a variety of research fields (e.g. biology, life sciences, ecology, social sciences). Both organizations organize short, domain-specific workshops of two to three days in which basic principles are taught by trained instructors. Both organizations adhere to a particular teaching style that values hands-on programming by participants, and live coding of instructors. Whereas Data Carpentry is primarily directed at learners without prior programming experience, Software Carpentry workshops often require a basic knowledge of programming languages such as R or Python.

Besides focused workshops, the challenges of data-driven science may ask for more formal training in terms of dedicated Master's programs. Such Master's programs could provide specialized training at the cross-section of ecology and computer science, covering advanced topics such as database creation and management, computer vision and machine learning algorithms, geographical information systems (GIS) and modeling of complex ecological systems.

We concluded the plenary discussion with agreement that both disciplines would benefit from a better dialogue.

---

[1] https://github.com/opetchey/RREEBES
[2] https://software-carpentry.org/
[3] http://www.datacarpentry.org/

References

1. Freckleton, R. P. and Iossa, G. (2010). *Methods in ecology and evolution.* Methods in Ecology and Evolution, 1(1):1–2.
2. Michener, W. K. and Jones, M. B. (2012). *Ecoinformatics: supporting ecology as a data-intensive science.* Trends in Ecology & Evolution, 27(2):85–93.
3. Mislan, K. a. S., Heer, J. M., and White, E. P. (2016). *Elevating The Status of Code in Ecology.* Trends in Ecology & Evolution, 31(1):4–7.
4. Peng, R. D. (2011). *Reproducible Research in Computational Science.* Science (New York, N.y.), 334(6060):1226–1227.
5. *Development Core Team (2016).* R: A language and environment for statistical computing.
6. Roche, D. G., Kruuk, L. E. B., Lanfear, R., and Binning, S. A. (2015). *Public Data Archiving in Ecology and Evolution: How Well Are We Doing?* PLOS Biology, 13(11):e1002295.
7. Sandve, G. K., Nekrutenko, A., Taylor, J., and Hovig, E. (2013). *Ten Simple Rules for Reproducible Computational Research.* PLOS Computational Biology, 9(10):e1003285.
8. Shade, A. and Teal, T. K. (2015). *Computing Workflows for Biologists: A Roadmap.* PLOS Biology, 13(11):e1002303.
9. Whitlock, M., McPeek, M., Rausher, M., Rieseberg, L., and Moore, A. (2010). *Data Archiving.* The American Naturalist, 175(2):145–146.
10. Whitlock, M. C. (2011). *Data archiving in ecology and evolution: best practices.* Trends in Ecology & Evolution, 26(2):61–65.
11. Wilson, G., Aruliah, D. A., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., Haddock, S. H. D., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B., White, E. P., and Wilson, P. (2014). *Best Practices for Scientific Computing.* PLOS Biology, 12(1):e1001745.

Participants

- Martin Bücker
Universität Jena, DE
- Tilo Burghardt
University of Bristol, GB
- Gustau Camps-Valls
Universitat de València, ES
- Yun-Heh Jessica Chen-Burger
Heriot-Watt University – Edinburgh, GB
- Joachim Denzler
Universität Jena, DE
- Matthew Evans
University of Hong Kong, HK
- Florian Hartig
Universität Regensburg, DE
- Thomas Hickler
Senckenberg Research Centre, DE
- Donald Hobern
GBIF – Copenhagen, DK
- Forrest Hoffman
Oak Ridge National Laboratory, US
- Kazuhito Ichii
JAMSTEC – Yokohama, JP
- Martin Jung
MPI für Biogeochemistry – Jena, DE
- Birgitta König-Ries
Universität Jena, DE
- Ivaylo Kostadinov
Jacobs University Bremen, DE
- Bertram Ludäscher
University of Illinois at Urbana-Champaign, US
- Miguel Mahecha
MPI für Biogeochemistry – Jena, DE
- Laetitia Navarro
iDiv – Leipzig, DE
- Shawn Newsam
University of California – Merced, US
- Frank Pennekamp
Universität Zürich, CH
- Natalia Petrovskaya
University of Birmingham, GB
- Markus Reichstein
MPI für Biogeochemistry – Jena, DE
- Andrew Richardson
Harvard University – Cambridge, US
- Ribana Roscher
Universität Bonn, DE
- Brody Sandel
Santa Clara University, US
- Bernhard Seeger
Universität Marburg, DE
- Johann Wolfgang Wägele
ZFMK – Bonn, DE
- Jakob Zscheischler
ETH Zürich, CH

Report from Dagstuhl Seminar 17092

Search as Learning

Edited by
Kevyn Collins-Thompson[1], Preben Hansen[2], and Claudia Hauff[3]

1 **University of Michigan – Ann Arbor, US,** kevynct@umich.edu
2 **Stockholm University, SE,** preben@dsv.su.se
3 **TU Delft, NL,** c.hauff@tudelft.nl

Abstract

This report describes the program and the results of Dagstuhl Seminar 17092 "Search as Learning", which brought together 26 researchers from diverse research backgrounds. The motivation for the seminar stems from the fact that modern Web search engines are largely engineered and optimized to fulfill lookup tasks instead of complex search tasks. The latter though are an essential component of information discovery and learning. The 3-day seminar started with four perspective talks, providing four different views on the topic of search as learning: interactive information retrieval (IR), psychology, education and system-oriented IR. The remainder of the seminar centered around breakout groups leading to new views on the challenges and issues in search as learning, interspersed with research spotlight talks.

Seminar February 26–1, 2017 – http://www.dagstuhl.de/17092
1998 ACM Subject Classification Information Systems; Information Retrieval; Users and Interactive Retrieval; Evaluation of Retrieval Results
Keywords and phrases Searching, Learning, Human information interaction, Search processes, Learning processes
Digital Object Identifier 10.4230/DagRep.7.2.135

1 Executive Summary

Claudia Hauff
Kevyn Collins-Thompson
Preben Hansen

Search is everywhere – it penetrates every aspect of our daily lives and most of us can hardly manage a few hours without resorting to a search engine for one task or another. Despite the success of existing (Web) search technology, there are still many challenges and problems that need to be addressed. Today's Web search engines (often also powering domain-specific and site-specific search) are engineered and optimized to fulfil individual users' lookup tasks. This efficiency, however, also means that we largely view search systems as tools to satisfy immediate information needs, instead of rich environments in which humans heavily interact with information content, and search engines act as intelligent dialogue systems, facilitating the communication between users and content. Web search engines are not designed for complex search tasks that require exploration and learning, user collaborations and involve different information seeking stages and search strategies, despite the fact that more than a quarter of Web searches are complex. In recent years, there has been a growing recognition of the importance of studying and designing search systems to foster discovery and enhance the learning experience during the search process outside of formal educational settings. Searches that lead to learning, are naturally complex. Research progress in this area, however, is slow, with many more open questions than answers. Several critical bottlenecks and major impediments to advancements in the search as learning area exist, including (i) the reliance on small-scale lab studies to evaluate novel approaches which severely limit the diversity of investigable factors as well as the ecological validity and generalizability of the findings; (ii) the lack of awareness among researchers' initiatives in this very multidisciplinary area of work; and (iii) the lack of a shared research infrastructure. The 3-day seminar gathered 26 prominent researchers from the fields of information retrieval, psychology and the learning sciences in order to address the critical bottlenecks around search as learning. The seminar sessions alternated between tutorial-style presentations to learn from each other's disciplines and interactive breakout sessions to find a common ground and address the most pressing issues related to the four big research themes of (i) understanding search as a human learning process; (ii) the measurement of learning performance and learning outcomes during search; (iii) the relationship between the learning process and the search context; and (iv) the design of functionalities and search system interventions to promote learning.

2 Table of Contents

3 Overview of Talks

3.1 What do learners gain when searching?

Leif Azzopardi (University of Strathclyde – Glasgow, GB)

When searching a person faces many choices and decisions e.g. what to query for, what to examine, when to stop? We have developed many conceptual and descriptive models that describe how people interact and what factors are likely to influence the choices that people make. However, they lack the power to predict how a person will behave or explain how they act. This is where formal, often mathematical, models come into play – and have been reasonably successful in explaining how people search and why they behave they way they do, e.g. Foraging Theory and Economic Theory. The basis of such models, requires the specification of a cost function and a gain function, from which we can determine the optimal search behaviour. I hypothesize that one's search behaviour and performance is indicative of their expertise and how close they are to optimal – and so it would be interesting to evaluate and assess the search behaviours and performance of learners to determine whether this is the case or not. However, this is all premised on the measuring of costs and gain. So what do people, in particular, learners, gain when they search? I argue, that we need to move beyond the notion of binary relevance to modelling and measuring the usefulness/utility/value of the information encountered during the search process.

3.2 Crowdsourcing

Ujwal Gadiraju (Leibniz Universität Hannover, DE)

Crowdsourcing has become an increasingly popular means to acquire human input on demand. Microtask crowdsourcing marketplaces facilitate the access to millions of people (called workers) who are willing to participate in tasks in return for monetary rewards or other forms of compensation. This paradigm presents a unique learning context where workers have to learn to complete tasks on-the-fly by applying their learning immediately through the course of batches of tasks. However, most workers typically drop out early in large batches of tasks, depriving themselves of the opportunity to learn on-the-fly through the course of batch completion. By doing so workers squander a potential chance at improving their performance and completing tasks effectively. In this talk, we propose a novel method to engage and retain workers, to improve their learning in crowdsourced information finding tasks by using achievement priming. We present rigorous experimental findings that show that it is possible to retain workers in long batches of tasks by triggering their inherent motivation to achieve and excel. As a consequence of increased worker retention, we find that workers learn to perform more effectively, depicting relatively more stable accuracy and lower task completion times in comparison to workers who drop out early.

3.3 The user's emotional experience in learning and search processes

Gabriele Irle (Universität Hildesheim, DE)

The central question that motivates my research is, how searchers emotionally experience online information searches and which causes and conditions they consider as significant for their emotional experience during the search. Firstly, the results demonstrate that the individual search topics have a strong influence on the searchers' feelings. Secondly, it is shown that those feelings that refer to the activities within the search process are surprisingly less pronounced, because the online search is considered a routine activity. As a result of these findings, we discussed if the search process is secondary to learners, while they are primarily interested in the subject matter itself. For future research on search as learning, we suggest to differentiate between process emotions, prospective and retrospective emotions as well as social emotions.

3.4 Search as learning – a psychological perspective

Yvonne Kammerer (IWM – Tübingen, DE)

In my talk I gave an overview on the topic "search as learning" from a psychological perspective (specifically an educational and applied cognitive psychology perspective). The focus of psychological research in this field is on using the Internet to learn about complex, conflicting scientific or health-related issues rather than to learn simple facts. Such so-called ill-structured problems do not have a single, definitive solution, but are characterized by conflicting and fragile evidence. Two central processing steps that are typically addressed in this mostly experimental research are (1) the evaluation and selection of search results presented by a search engine, and (2) the comparison and integration of information from multiple websites. Moreover, during both steps source evaluation processes are investigated; i.e., whether, how, and when learners attend to, evaluate, and use information about the sources of documents (cf. credibility assessment). As outlined in my talk, a central goal of psychological research in this field is also to identify and examine factors that might influence the information seeking processes and learning outcomes. Such influencing factors are, for instance, prior topic knowledge or attitudes (i.e., individual variables), task instructions or trainings (i.e., contextual variables), or search tools or interfaces (i.e., resource variables).

3.5 How can SAL studies help search engines

Yiqun Liu (Tsinghua University – Beijing, CN)

As the information needs of Web search engine users become more and more diverse, complex search activities, such as exploratory search and multi-step search, have been identified and considered challenging for current search systems. As the user plays a central role in the highly interactive complex search session, user behavior analysis and modeling is vital for making search engines more effective for learning oriented search tasks. After looking through existing researches in search as learning studies, we believe that three research questions should be focused on: (1) How to model search users' cognitive states. (2) How do users' cognitive states affect search behaviors. (3) What are the implications for search engines. Especially, we believe that the third question is quite important because it is highly related with how we can improve current commercial search engines to support SAL processes.

3.6 SAL – A Information Retrieval (IR) / Interactive Information Retrieval (IIR) perspective

Heather O'Brien (University of British Columbia – Vancouver, CA)

Within interactive information retrieval (IIR), learning has been defined as the acquisition of information for the purpose of changing, augmenting, or reinforcing one's existing knowledge base. Learning is intertwined in the search process: searchers learn through interacting with search tools and multimedia content, and is associated with search tasks that require interpreting, comprehending, and comparing information. Within IIR, "searching as learning" is a burgeoning area, but there are many challenges that need to be addressed. First, there is the issue of measurement in dynamic search contexts. Specifically, there is the question of how to develop and evaluate the utility and robustness of learning measures, and the need to distinguish short- and long-term learning outcomes. Second, designing search systems that support learning is essential, and work in the area of exploratory search has addressed the need for systems to support learning, engagement and discovery. However, these systems must not only allow people to explore information, but must support cognitive and affective learning needs. For example, imagine systems that help searchers save, annotate, and revisit information to manage their cognitive load, or that re-engage them if they become frustrated or confused during the learning process. Lastly, we know little about searchers as learners, and how their goals and cognitive abilities interact with system and content variables to influence learning outcomes. Future work must focus on these and other challenges in order to deepen our theoretical, methodological and applied knowledge and contribution in this area.

3.7 Searching As Learning

Rebecca B. Reynolds (Rutgers University – New Brunswick, US)

My work addresses people's learning with information systems, situated in social constructivist learning theory. In conceptualizing and conducting my research I begin by considering the human actor. I investigate ways in which people engage with one another and human-produced documents, artifacts and products, as they construct and produce ideas, knowledge, understanding and artifacts. Social constructivist learning interventions come in many shapes and sizes; my work considers how the structure of such contexts shape and contribute to learning processes and outcomes. My work is situated in information science as well as the learning sciences. In the learning sciences, scholarly debates have emerged around the effectiveness of inquiry-based learning approaches such as those involving autonomous information-seeking and resource uses, on account of cognitive load (e.g., Kirschner, Sweller, Clark, 2006; Hmelo-Silver, Duncan, Chinn, 2007). My work addresses these issues, as I consider ways to develop instructional affordances that optimize learning potentials in users.

Blended e-learning is becoming more commonplace in K-12 education – those contexts in which students meet face-to-face in classrooms, while also using social and informational affordances made available in online environments such as learning management systems and search engines. In one line of work, I investigate middle school and high school student uses of blended learning affordances, specifically in the domain of computer science education. I study an educational project involving students' design and programming of interactive web games. Students and teachers in this context engage daily in a formal in-school class, using a blended learning information system containing the curriculum: multi-modal tutorials, information resources, sequenced assignments to complete online, social engagement affordances, code libraries and presentation spaces for middle schoolers' work, online social features, etc. I investigate children's motivational and information-seeking dispositions and processes, as well as learning outcomes in this context (Reynolds & Harel, 2011; Reynolds & Chiu 2012, 2013, 2015). My work considers how specific instructional features of a blended e-learning curriculum can be optimized to improve students' information-seeking and knowledge-building processes and outcomes (Reynolds, 2016b). In doing so, I consider the role of a range of resources in the learning context ecology, which in addition to the online system, includes expert peers and teachers in class (Reynolds 2016a). My findings indicate that students vary in their need for structure (2012, 2013, 2016a); a robust blended learning environment for young people's CS education can meet student needs by offering multi-modal resource availability, and when led by an educator with sufficient expertise.

My work also considers generalizability of my findings, to other settings, disciplinary subject domains, and/or user populations (e.g., Chu, Reynolds et al., 2016). I am also exploring reading and literacy levels as a factor in student processing of online informational texts for their productive task completion – like that of designing a game.

References

1. Chu, S., Reynolds, R., Notari, M., & Lee, C. (2016). In: Chu, S., Reynolds, R., Notari, M., Taveres, N., & Lee, C. 21st Century Skills Development through Inquiry Based Learning: From Theory to Practice. Springer Science. 214 pages.
2. Hmelo-Silver, CE., Duncan, RG., & Chinn, CA. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). Educational Psychologist, 42, 99–107.
3. Kirschner, PA., Sweller, J., & Clark, RE. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41, 75–86.
4. Reynolds, R. (2016a). Defining, designing for, and measuring "digital literacy" development in learners: A proposed framework. Educational Technology Research & Development. 64(1).
5. Reynolds, R. (2016b). Relationships among tasks, collaborative inquiry processes, inquiry resolutions, and knowledge outcomes in adolescents during guided discovery-based game design in school. Journal of Information Science: Special Issue on Searching as Learning. 42: 35–58.

6 Reynolds, R. & M. M. Chiu. (2015). Reducing digital divide effects through student engagement in coordinated game design, online resource uses, and social computing activities in school. Journal of the Association for Information Science and Technology. DOI: 10.1002/asi.23504
7 Reynolds, R., & Chiu, M. (2013). Formal and informal context factors as contributors to student engagement in a guided discovery-based program of game design learning. Journal of Learning, Media & Technology. 38(4), pp. 429–462.
8 Reynolds, R., & Chiu, M. (2012). Contribution of motivational orientations to student outcomes in a discovery-based program of game design learning. Proceedings of the International Conference of the Learning Sciences (ICLS), July 2012, Sydney, Australia.
9 Reynolds, R., & Harel Caperton, I. (2011). Contrasts in student engagement, meaning-making, dislikes, and challenges in a discovery-based program of game design learning. Educational Technology Research and Development, 59(2), pp. 267–289.

3.8 Information behavior in educational information systems – Teachers searching for lesson preparation

Marc Rittberger (DIPF – Frankfurt am Main, DE)

Searching is one of the main competences with respect to the digitalization of education. In Germany the "Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (KMK)" states in her strategy "education in the digital world" that searching, processing and curating are main competences needed in education. We think, that one of the consequences for learning may be, that education, e.g. in schools, will change from knowing about things to also know where and HOW to find things. In this context of digitalization the organization of learning environments is changing, e.g. by using open educational resources. In our research we observed teachers searching for lesson preparation and we analyzed several data resources, where teachers are searching for learning materials. Results show, that teachers in German speaking countries demand quite concrete questions in the fields of content, method, or aims of lessons. Questions are less concrete, if teachers have information needs with respect to control, sanctions, or organization of the lesson.

3.9 Search as Learning or Learning by Search

Marcus Specht (Open University – Heerlen, NL)

Searching information is a key component of nearly all human learning processes. Recently the discussion about information literacy has become very popular especially when using web search engines and digital information repositories in educational settings. At the Welten Institute we have been exploring different directions for supporting teachers and learners in search activities. Desktop research is a major search activity being used from primary school to university level. The major elements of search activities include the development of expertise to define the right search terms, evaluate available sources and documents, information integration of different sources and other activities depending on the didactical setting. In the WI we perform empirical research on the development of expertise and information skills and develop models for information skills and optimal design of supporting ICT tooling and instructional design.

For the development of expertise the understanding of a domain structure and the different taxonomies used in the domain are a major difference between experts and novices. This is the starting point for a set of projects where our researchers have been working on the use of taxonomies and classification systems for supporting learners in the use of digital information repositories. The WI has developed different visualisations, tools and technologies for personal and collaborative exploration of big data and information spaces focusing on learning and developing expertise.

In designing tools, technologies and instruction to enable humans to search and explore big data structures and use them in learning the WI contributes to the current and future scenarios for Search As Learning.

3.10 Computational Metacognition

Michael Twidale (University of Illinois at Urbana-Champaign, US)

I am trying to understand the ways that people who seem good at learning technologies (often self-identified as 'techies') go about doing that learning, and how that contrasts with people who seem less good at learning technologies (often self-identified as 'non-techies'). Much of the difference seems to be due to a range of metacognitive strategies, combined with certain attitudes to technology learning and the absence of certain misconceptions about technology learning. A critical set of Computational Metacognitive strategies seem to revolve around techniques for searching for information, help and insight that can further the desired technology learning and overcome problems and confusions. These techniques help address challenges of how to search for something when you don't know what it is called, how to assess the quality of results and their usability, usefulness and appropriateness for one's level of understanding of the domain (complete but bewilderingly complex answers may be worse than superficial, limited but comprehensible tech fixes), how to handle complex goalstacks of levels of prerequisite actions and knowledge, and how to tailor partial solutions to address the actual problem at hand. The goal is to design pedagogies, interfaces and functionality to enable a much broader proportion of the population acquire the skills to better manage their own learning of technology.

3.11 Looking for "Listening" Online: A Learning Sciences & Learning Analytics Research Project with Potential Implications for Studying & Supporting Search as Learning

Alyssa Wise (New York University, US)

This presentation overviews the notion of online listening as a vehicle for examining how learners attend to the contributions of others in online spaces; this is part of an effort to broaden our understanding of the process of learning through dialogue (with others, potentially also with resources) in technology-mediated environments. Online listening is defined as the collection of behaviors learners engage in as they interact with the existing posts in a discussion forum; it is specifically differentiated from the act of lurking (which implies a lack of subsequent contribution).

Drawing on five years of research conducted through the E-Listening research project, the presentation includes a (re)conceptualization of the notion of listening for online spaces; an explanation of a theoretical taxonomy for considering different kinds of listening in online discussion; documentation of empirical findings about both the specific patterns of listening in which students engage and their the relationships of these to subsequent contributions; the design of a novel graphical interface to support more productive online listening behaviors; and finally the provision of learning analytics on online listening activities to students to support their self-regulation of such. Collectively, results indicate that online listening is a useful concept for investigating the ways learners interact with existing posts in online discussions and for designing technological and pedagogical interventions to support more productive participation. Parallels to notions of search and learning and the potential implications for studying and supporting this process are discussed.

3.12 Dynamic Information Retrieval Modeling

Grace Hui Yang (Georgetown University – Washington, US)

Many modern IR systems and data exhibit characteristics which are largely ignored by conventional techniques. What is missing is an ability for the model to change over time and be responsive to stimulus. Documents, relevance, users and tasks all exhibit dynamic behavior that is captured in big data sets (typically collected over long time spans) and models need to respond to these changes. Further to this, advances in IR interface, personalization and ad display demand models that can react to users in real time and in an intelligent, contextual way. This talk provides an introduction to Dynamic Information Retrieval Modeling. In particular, I talk about how we model information seeking as a partially observable Markov decision process. I also talk about the TREC Dynamic Domain Track.

3.13 Challenges in Measuring Knowledge State Change during Search Sessions

Ran Yu (L3S Research Center – Hannover, DE)

Search systems to date have been viewed more as tools for the retrieval of content to satisfy information needs, than as environments in which humans interact with information content in order to learn. There are more and more research works focusing on improving the learning experience and efficiency during search sessions, however, the measurement of a user's knowledge gain is a common challenge that has not yet been addressed.

In the SoA works, in order to evaluate their approach, the researchers limited their experiment to a few (usually 1-3) very specific predefined topics. Afterwards, they measure the knowledge gain by conducting a small scale quiz or questionnaire. This has limited the scope as well as the contribution of the works.

In a recent project, we focus on measuring knowledge gain during learning related search sessions based on the search activities. This can potentially provide a way to conduct real time evaluation of learning performance without requiring users to provide extra information. This can benefit the works for SAL optimization. This talk is about the challenges that we have encountered in this project, which includes the identification of learning activities, the modeling of knowledge gain and the lack of open data.

4 Working groups

4.1 Working Group Summaries

Kevyn Collins-Thompson (University of Michigan – Ann Arbor, US), Preben Hansen (Stockholm University, SE), and Claudia Hauff (TU Delft, NL)

The following summaries were produced from all working groups. Each group's discussion revolved around a central question, stated at the top of the summary. At the end of each discussion, the group listed challenges and opportunities for future work.

4.2 Understanding search as a human learning process: when and how does learning occur in the search process?

Participants: Heather O'Brien, Christa Womser-Hacker, Soo Young Rieh, Alyssa Wise, Gabriele Irle, Ran Yu, Dan Russell, Rebecca Reynolds, Claudia Hauff.

Introduction

Few empirical studies exist (e.g. [2, 3, 1]) that attempt to quantify when and to what extent learning occurs during the search process. These studies were a starting point for our discussion as they showcase some of the challenges we face in our quest to understand when and how learning occurs in the search process. As this breakout group was the first of the seminar, it covered a wide range of topics and identified a wide range of challenges, not all of them strongly connected to the initial question.

Challenges

1. Different types of learning (procedural vs. declarative) and different learning contexts (structured vs. unstructured, formal vs. informal, incidental vs. intentional) exist. Investigating the impact of search in these circumstances and disentangling those is likely to require different evaluation metrics that are designed specifically for each type and context. It is an open question whether the currently employed proxies (such as measuring the broadening of the vocabulary used in the searches) is a sufficiently clean signal to measure learning.
2. Is it possible to measure to what extent robust learning (learners achieving a deep conceptual understanding) or transfer learning (learners employing learnt concepts in novel situations) is taking place during the search process?
3. Search query logs offer a very limited view into users' minds; we have to make educated guesses on their learning intent, their prior expertise and their context based on noisy signals. In order to make strides into understanding learning we require large-scale data with more semantic meaning behind it. How would such data look like and how can we collect it at scale? This challenge also ties in with the question whether search should be at the centre of the investigation or 'just' one block in the ecology of learning?
4. How can we help users that want to learn something but already struggle early on in the search process when formulating an initial query based on their information need? A common use case here are medical inquiries with users querying for symptoms in laymen terms ("pain in my side") .
5. In the formal learning setting the instructor plays an important role. Does the instructor also have a role in the search as learning setting and if so, how can that role be supported algorithmically or on the search interface level?
6. In formal learning, scaffolding (the learning material is broken down into pieces and a structure is imposed on each piece) is an important part of lesson design. Is it possible to design automated scaffolding tools into the search process? While one could consider query autocompletion and query suggestions as already existing scaffolds, they do not provide sufficient guidance to the learner.
7. What impact does evolving knowledge (Is Pluto a planet? How many moons does Saturn have?) have on search as learning? How can we support users that are searching for information with search requests that are already providing a certain answer frame (e.g. "vaccinations are bad", "climate change is not real")?

References

1 Eickhoff, Carsten, Jaime Teevan, Ryen White, and Susan Dumais. "Lessons from the journey: a query log analysis of within-session learning." In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, pp. 223–232. ACM, 2014.

2 Jansen, Bernard J., Danielle Booth, and Brian Smith. "Using the taxonomy of cognitive learning to model online searching." *Information Processing & Management* 45, no. 6 (2009): 643–663.

3 Wilson, Mathew J., and Max L. Wilson. "A comparison of techniques for measuring sensemaking and learning within participant-generated summaries." *Journal of the American Society for Information Science and Technology* 64, no. 2 (2013): 291–306.

4.3 What is the relationship between the learning process and the context (educational, work-related, etc.) in which learning & searches occur?

Participants: Michael Twidale, Yiqun Liu, Ujwal Gadiraju, Grace Hui Yang, Marc Rittberger, Marcus Specht, Dirk Ahlers, Kevyn Collins-Thompson.

Introduction

Understanding the context of information requests can be critical to interpreting and satisfying them correctly, particularly to support learning. This discussion explored the following themes, beginning with how to characterize and identify "context" in the first place, to methods for getting a more complete picture of a learner's contexts, and the relationship between context and learning.

Issues

How to characterize the dimensions of context. Those mentioned included location, time, task and workflow stage, intention, user characteristics (role/persona, background knowledge, demographics like age, prior knowledge, user history), environmental/machine/resources available, relationships with people and organizations – and the basic questions who, what, where, when, why, and how. Also discussed was the idea of developing a shared open taxonomy of context definitions, for research study definition, evaluation, and application purposes.

Getting a more complete, continuous picture of context for learning. A recurring theme was the problem that currently, search algorithms see only a very limited amount of contextual information (e.g. during the time a user interacts with an online system, such as previous queries in a session) which might limit their ability to find the right information for a learning need. For example, would it help to know more about what a person did before or after they went online? How could a system obtain a more complete picture of a user's continuous, multi-faceted context throughout the day, and connect these with the many different tasks per day? Another key problem in this area is how to tease apart multiple overlapping types of context. There was disagreement as to whether priority should be given to explore the promise of new sensors or signals to fill in these gaps, versus focusing on better processing and integration of existing signals that we already have. The question was also raised: how much context is actually necessary for each task? Can we get context from mental state, and distinguish conscious vs subconscious intent? One suggestion was that an important case of informative contexts were those requiring deep understanding of higher-level scenarios, such as stressful situations, or social atmosphere, where a little information can go a long way.

The relationship between context and learning. Participants noted the importance, and challenge, of matching the learning task to the appropriate context(s), and identifying which contexts were more or less supportive of learning. Sometimes, there is no explicit or conscious learning goal, and content might help identify these cases. A recurring theme was the need to identify the desired learning outcomes that are associated with various activities in context. In fact, the actual learning might happen later, after a user leaves a particular context. Related to theories of situated learning, and the problem of defining the context/situation that something is relevant, we noted as an example of work in learning/context interaction, and cross-context learning, work such as that described in the STELLAR RTST Trend Report on Contextualisation [3], e.g. with applications in language learning: distributed scenarios (integrate situations, devices). Other related work included personalized academic search [2] and information access by professionals in workplace settings [1].

Opportunities

The discussion ended with participants answering a hypothetical question of what research direction they would explore in this area if given a large research award. Some ideas mentioned included the following:

1. Tracing and measuring learning trajectories in everyday life: a longitudinal understanding of cross-contextual, informal, and unintentional learning.
2. How can we build interfaces to externalize models/thinking, including of benefit to yourself, to get the effect of UI where people benefit from reflection, get it out of their own heads, personalize, and in a way that promotes collaboration?
3. A more complete picture of context: follow up search trails across different silos, study what happens offline, filling gaps between online accesses.
4. Search within an automotive context: use cameras, multisensors to provide context.
5. Contextual mobile search: use sensors to determine when and how to search.

Challenges

1. How can we get at contextual gaps that we can't currently observe?
   a. Resources include: ethnographers, log analysis, hardware/Kinect/sensor engineering, HoloLens, motion sensors, full body sensor, Amazon Echo, observing and capturing the work of experts.
   b. Use existing tools to look at awkward and/or more detailed data.
2. How can algorithms know what signals matter in context?
3. How can we tease apart multiple contexts that happen over time?
4. Does generating context help with reflection, meta-cognition, and learning?
5. How can we identify and exploit incidental learning opportunities in context?
6. How does context interact with learning? How can we measure learning in all of these different contexts?
7. How can we organize the context types we have, and how can we understand cost-benefit tradeoffs, and determine which features are most important (to measure, to infer, or to ask people about)?
8. How can we deal with context latency, where it takes time to process key features of a context, which may then become "stale" or less relevant?
9. How can we exploit context without it exploiting us: privacy issues?

References

1 Ahlers, D., Mehrpoor, M., Kristensen, K., Krogstie, J. Challenges for Information Access in Multi-Disciplinary Product Design and Engineering Settings,. ICDIM 2015. DOI: 10.1109/ICDIM.2015.7381865

2 Suzan Verberne, Maya Sappelli, Diana R. Sørensen, Wessel Kraaij (2013). Personalization in Professional Academic Search. In: M. Lupu, M. Salampasis, N. Fuhr, A. Hanbury, B. Larsen, H. Strindberg (eds.): Proceedings of the Integrating IR technologies for Professional Search Workshop at ECIR, Moscow, Russia, 24 March 2013.

3 Specht, M., Börner, D., Tabuenca, B., Ternier, S., De Vries, F., Kalz, M., Drachsler, H., & Schmitz, B. (2012). RTST Trend Report: lead theme Contextualisation. Deliverable 1.7 of STELLAR network of excellence. Heerlen, The Netherlands. http://dspace.ou.nl/handle/1820/4356

4.4 How can constructs and results from cognitive psychology and education be used to inform models of learning, knowledge acquisition, or mental representation during search interaction?

Participants: Gwen Frishkoff, Tim Gollub, Andreas Nürnberger, Yvonne Kammerer, Leif Azzopardi, Jiyin He, Rob Capra, Noriko Kando, Preben Hansen.

Introduction

The breakout group started with a general introduction and discussion of some of the general and basic models and frameworks used in cognitive psychology, education and IR. Especially, classical algorithmic IR and the interactive IR model, such as the Marcia Bates berry-picking model, Carol Kuhlthaus' ISP model and Dervin's sense making models, but also Pirolli's Information Foraging Theory [4] were discussed.

In interactive IR an episode could range from seconds in duration, to days or weeks, depending on the domain (for example, within the patent domain).. When working in the area of neuro-cognition, an episode usually lasts a second or less. The fMRI, EEG and ERP techniques were discussed and suggestions how they could be used were proposed. Temporal patterns and semantic priming can then be detected. The group also discussed the differences between performing laboratory tests and studies in natural setting and real-time.

For example, on the question if you want to measure if a person has understood something in a learning situation, do you need to do some prior measures? The answer to this question, suggested that a path to take could be to use measures of semantic priming. Another example is if a person is searching for a certain concept and then find it, this activity could be measured with biomarker comprehension. There is no cognition without motivation and emotion is part of the cognition. Domain general processes are learning and mental processes, while domain specific processes are processing things like sound and images.

From the cognitive psychology IS/IR area we can use several benefits of implicit methods, such as EEG/eye tracking (i.e. measures that do not rely on explicit responses to questions about learning, search outcomes). Vocabulary learning from multiple contexts illustrates incremental learning of information over multiple, diverse instances. The amount of information that the mind/brain is processing at any one time is large and mostly implicit (unconscious). Only a small fraction reaches consciousness, and conscious reflections on learning may or may not be correct. However, from a cognitive psychology point of view, the traditional search task descriptions are too complex and difficult to utilize.

Challenges

1. Dealing with information searching as learning, how studies and different data collection methods could be operationalized when an episode of a study could be a second or lesser?
2. Information Seeking and Information Retrieval use and apply many different kinds of methods. What can we use from cognitive psychology to correlate these methods? There are however, several challenges in using implicit measures: need for validation, cross-validation methods.
3. A third challenge mentioned is how to use ERP in a natural setting.
4. What if we could instrument every single second during a research study. How could it be utilized within the research of interactive IR? For example, if it is important to see a person spending effort in a certain learning situation and we want to measure it. One example discussed was to measure frustration.
5. How can different biomarkers be used? Temporal and stable markers is also something that can be measured and used to decompose obtained patterns.

References

1 Low, T., Bubalo, N., Gossen, T., Kotzyba, M., Brechmann, A., Huckauf, A. and Nürnberger, A., Towards Identifying User Intentions in Exploratory Search using Gaze and Pupil Tracking, Proc. of ACM SIGIR Conference on Human Information Interaction & Retrieval (CHIIR) (2017). http://sigir.org/chiir2017/accepted-papers.html
2 Mostafa, J., Carrasco, V., Foster, C., & Giovenallo, K. (2015). Identifying Neurological Patterns Associated with Information Seeking: A Pilot fMRI Study. In Information Systems and Neuroscience (pp. 167–173). Springer International Publishing.
3 Mostafa, J., & Gwizdka, J. (2016, March). Deepening the Role of the User: Neuro-Physiological Evidence as a Basis for Studying and Improving Search. In Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (pp. 63–70). ACM.
4 Pirolli, P. (2007). Information Foraging Theory. Adaptive interaction with information. New York: Oxford University Press.

4.5 What type of functionalities and interventions on the search system interface level and the search algorithm level can foster learning?

Participants: Dan Russell, Preben Hansen, Michael Twidale, Gabriele Irle, Tim Gollub, Rob Capra, Soo Young Rieh, Grace Hui Yang, Dirk Ahlers, Yvonne Kammerer, Marcus Specht, Christa Womser-Hacker, Kevyn Collins-Thompson.

Introduction

This discussion focused not only on specific user interface affordances that could help learning, but also on issues and design principles that are important to keep in mind when creating such affordances for learning during search.

Issues

The goals and effects of user interface additions. We began this discussion with one participant describing a commercial search engine company's experiment to study the effect of new additions to the user interface. For example, some tooltips helped with advanced query operators, but some made things worse, because people would misinterpret instructions. So any interface design must be done with sensitivity that users are very different than search developers: we need much more information about a user. In general, a lot of search functionality is developed that doesn't invest time in this – new features are just deployed to test online, at scale. Other examples of consequences associated with new interface additions included dialog systems, in which user perceptions are very sensitive to timing – in general, sociolinguistic behavior and cues need to be considered as part of how users will interpret information. The use of benevolent deception [2] was also discussed, including the use of models that are simple but not necessarily accurate (as one related example: query autocompletion models can be heavily edited for simplicity/popularity, with the major effect of that feature to speed up query creation, not necessarily to provide the best alternatives for that user and query).

Interface elements that foster learning. Google Squared [1] was mentioned as an exploratory tool that would allow a user to find different common aspects of a class of objects, e.g. mammals, by adding a column to those from an existing set of Google Squared results. This is similar to the ability to find analogies between items: one could explore the conceptual space. Dialog-based systems have potential to help learning: recall that librarians used to ask followup questions. One advantage of agent-based systems is that they can get more context e.g. also through followup questions that go beyond the topic, or even the prosody of the conversation. Finally, allowing users to specify how they want to make tradeoffs might help them learn implicitly about the structure of a topic space (e.g. a taxonomy of objects) without explicitly showing them directly (leading to better retention). One example was mentioned from the Kayak travel website, which let users set sliders to specify tradeoffs instead of having to issue iterated queries – such features could give users a sense of sensitivity and immediate feedback, to understand how much narrowing down they need to do in a topic area.

How should new interfaces or algorithms be optimizing for better learning? As a counterpoint to "smart" search algorithms, the group discussed the idea that users sometimes just want to know what button to press – i.e. they just want an efficient, predictable tool. For learning, there is also a need to slow down, and be less efficient: time spent searching may be an investment in the future, and there may not be an immediate reward. Users may not always act with rational economic behavior, however. The use of conversational models has promise, but perhaps one goal should be to help students know what they don't know. Another goal could be to encourage broadening or lateral moves when exploring a topic: from a learning perspective, we often need a diversity of sources. This led to the question: what kinds of diversity are important for learning? In another direction, visualization approaches could also help users discover and explore, e.g. "Metro"-style maps of information [3].

Ideas for interface functionality

During the session, participants mentioned a number of ideas – from broad to very specific – of interface affordances or system functionality that could help with learning, or with developing learning-based systems.

1. Show a "controversy level" for search results, or flag when sources of points of view are not diverse enough.
2. Along the lines of search as a service (in particular, a dialog-based service): create a "search for learning" service to be shared. This could be an asynchronous service, so that search could deliver results over time in the background.
3. Patent search: add count info to query results to suggest where a patent examiner should focus.
4. Make it easier for humans to recall items and commands by providing cues (e.g. change in visual appearance).
5. Add adjunct elements to default search, e.g. through the addition of a sidebar. Users would start a search on their own but access the sidebar when stuck. For this, you would need to build in any new functionality during the transition period where a user learns to trust the new thing learns to use it effectively. This "persuasion" or "mentoring" itself would require new interface features. Users need to be able to control when they can invoke the new learning widget vs. resuming regular search.
6. Encourage users to type longer queries by using nudging behavior, e.g. a halo around text that would change color as users type.
7. Use taxonomies as a scaffold for certain kinds of learning tasks – perhaps via elastic lists.

Challenges and Opportunities

1. Increase dialog, followup questions, and more generally, the ability of a search system to engage in conversational agent behavior. More generally, we need a "context creation engine".
2. How can we track what users do with the information *after* the search to make sense of the information, and recognize the value of the search system?
3. How can we define diversity (in the search result sense) along different dimensions (topic, opinion, etc) and which types of diversity are appropriate for a given learning task?
4. What hints can we give users about search results and search process toward benefits for learning (e.g. diversity, serendipity, discovery, other users trails)?
5. Better exposure of the dimensionality of the retrieval space for user navigation.
6. When and how to target user learning during search tasks, or incentivize users to switch to learning – how could a system know what point in a task a student is at (e.g. a scientific process)?
7. How can search engine interfaces encourage self-reflection or comparison to other users (for calibration, motivation)?
8. What higher-level types of learning (e.g. finding analogies) could be supported by slower and/or human-in-the-loop processes?
9. Develop adjunct services that support learning and reflection, in addition to commercial search (and have components talk with each other).

References

1 Google Squared. Wikipedia: https://en.wikipedia.org/wiki/Google_Squared
2 E/ Adar, D.S. Tan, and J. Teevan. 2013. Benevolent deception in human computer interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'13). ACM, New York, NY, USA, 1863-1872. DOI: 10.1145/2470654.2466246.
3 D. Shahaf, C. Guestrin, E. Horvitz. Metro Maps of Science. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2012.

4.6 What are effective and scalable proxy signals for learning during search and search-related activities, and how can we measure and apply them?

Participants: Rob Capra, Yiqun Liu, Christa Womser-Hacker, Tim Gollub, Grace Hui Yang, Yvonne Kammerer, Andreas Nürnberger, Dirk Ahlers, Ujwal Gadiraju, Noriko Kando, Kevyn Collins-Thompson.

Introduction

In many cases, learning itself may be difficult, expensive, or invasive to measure directly. Proxy measures that are associated with learning (or lack of learning) can provide valuable indirect signals to information systems [2]. The discussion revolved around particular challenges to be solved, as well as ideas where there are opportunities for further progress.

Challenges

1. Which proxies to use and rely on will depend on the particular search task, so characterizing that interaction is an important goal.

2. How can we derive learning traces from a user's search trail? What are the relations we can get from such interaction sequences?
3. How can we study and support time-consuming queries that involve multiple steps or subtasks over extended periods: needs related to learning important facts or skills for life, financial management, job growth. This is hard to do in a lab study, where typically we give subjects the tasks, for a short period of observation.
4. How can we identify groups of people with similar learning needs in the same domain?
5. How can we detect what users need when learning: there are ultra-high priority learning needs: medical life-or-death search vs leisure learning.
6. How could we define a new learning channel where kids know they can go to (and which could be a source of new instrumentation)?
7. How could we use patterns of progression of query terms to gauge change in expertise over time, from progression of difficulty over time?
8. Most search models assume a cost-effective goal. What are other objective functions – and how can a search engine switch between these needs? E.g. hobbies, distraction. Should users tell the search engine directly? How should search engines account for e.g. a user's hierarchy of needs.
9. Can we measure user happiness as part of the user engagement scale?
10. Individual differences and personality need to be accounted for, but more work is needed on how to detect and support this. How inquisitive are you? How much do you need to understand?
11. There are three difficult prediction problems that should be distinguished. Can we know (1) *what* users learned vs. (2) *when* it happened vs. (3) whether they did learn *anything*? What classes of learning moments could be detected? For example, building on work by Yang et al. [5] can we identify learners' Eureka moments, from log and/or content features? Can we make use of additional signals from new sources, e.g. Kinect, skin conductivity.
12. How can we establish correlations between learning proxies and outcome measures, so that we can eventually reduce our reliance on more expensive/invasive proxies?

Opportunities

1. Rich content representations can lead to a host of new proxy features (e.g. work on reading difficulty and search [1]).
2. We could combine cognitive load and user navigational traces to identify learning goals. A certain amount of desirable difficulty may help indicate learning, or at least a context that is supportive of learning.
3. How can we think more broadly about educational applications, not just Web search. For example, the role of video (YouTube) as one of the main ways children learn can't be overstated at present, and that's underexplored.
4. Writing is a promising source of evidence of learning. At a high level, we can look at the semantics of the base content. Could we evaluate the quality of a summary according to how well it teaches a concept to someone else (or to a computer)? At a low low-level there's a lot of potential information in keystrokes and timing, e.g. what gets deleted and replaced.
5. Memory-retention/re-finding behavior could be a useful proxy, especially if we measure the difficulty of the concept being searched and re-searched – is it re-found more?
6. There are further opportunities for eyetracking : the video in mobile devices isn't precise, but we don't necessarily need high resolution for some important scenarios. For example,

we can detect reading or scanning by looking at e.g. back traces. We may also be able to discriminate between younger vs. older users, which could help disambiguate user characteristics [4].

7. Collaboration is still an area with many opportunities to understand and support learning: searching in a team, technology scouting, patent searching. How can we exchange learning results when there is an explicit learning trace? This could be applied in class settings: students could exchange traces [3]. Early work on collaborative IR captured some of this.
8. Studies of learning during test-taking, where the student is in a time-limited situation with access to online resources, could be a fruitful scenario to explore.
9. Most searches are not for learning. Are there learning-related verticals we could identify? Perhaps comparison shopping? Medical queries are one form of search where a user may be trying to learn something quickly.
10. There are further avenues for rich representations of users and tasks that exploit use of physiological signals: eye-tracking, skin, motion and video, audio.
11. There are likely to be further gains from work processing implicit signals: mining data from search trails, log features, queries and content, as well as explicit signals: mine data from e.g. collaborative communication.
12. Establish a learning-dedicated search "channel" that builds on existing commercial search as a service.
13. Establish correlations between proxy measures, outcome variables.

References

1 K. Collins-Thompson, P. N. Bennett, R. W. White, S. de la Chica, D. Sontag. Personalizing Web Search Results by Reading Level. Proceedings of the Twentieth ACM International Conference on Information and Knowledge Management (CIKM 2011). Glasgow, Scotland. Oct. 2011.

2 K. Collins-Thompson, S-Y. Rieh, C. Haynes, R. Syed. Assessing Learning Outcomes in Web Search: A Comparison of Tasks and Query Strategies (Full paper). ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR 2016). Chapel Hill, NC. March 2016.

3 Nürnberger, A., Stange, D., Kotzyba, M., Professional Collaborative Information Seeking: On Traceability and Creative Sensemaking, In: Proc. of 1st Intl. KEYSTONE Conference (IKC 2015), LNCS 9398, Springer, pp. 1-16 (2015) http://link.springer.com/chapter/10.1007/978-3-319-27932-9_1

4 Gossen, T., Höbel, J., Nürnberger, A., Usability and Perception of Young Users and Adults on Targeted Web Search Engines. In: Proc. 5th Inform. Interaction in Context Symposium (IIIX'14), pp.18-27, ACM (2014)

5 Hui Yang, Jiyun Luo, Christopher Wing. Detecting the Eureka Effect in Complex Search (short paper). In Proceedings of The 37th European Conference on Information Retrieval (ECIR 2015). Vienna, Austria.

4.7 How can learning performance and learning outcomes be measured during search? What search process features can act as indicators of learning?

Participants: Heather O'Brien, Ran Yu, Andreas Nürnberger, Noriko Kando, Yiqun Liu, Gwen Frishkoff, Ujwal Gadiraju, Leif Azzopardi, Jiyin He, Claudia Hauff.

Introduction

One of the most often recurring issues during the seminar was the issue of measuring learning gains that have occurred during search episodes. The standard approach of a pre-test → search episode(s) → post-test to measure these learning gains is not scalable, time-intensive to set up and only viable in a laboratory study. These inherent bottlenecks have led researchers to consider proxy measures (e.g. the change in vocabulary use during searches, the application of knowledge in downstream tasks) that are easier to collect at scale. A large part of the discussion revolved around scalability, different types of proxies and the ability to generalize beyond specific tasks and contexts.

Open questions

1. What scalable measures, based on search behaviours and document characteristics are good approximators of learning gains (and when is "good" good enough)? To what extent are measures task- and domain-specific? Across which periods of time (if we think for instance about sequences of queries across sessions) can we reliably measure learning gains?
2. Should the emphasis be on measuring learning, given all of the potential confounds? Or should we be looking at capacity to learn in the search process? There is also the role of probes in the search experience to test understanding as people make progress (challenge: level of intrusiveness).
3. Can we measure learning gains in downstream tasks, for instance by connecting users' search logs with their GitHub traces and observing the coevolution of both along several dimensions?
4. Laboratory studies have often elaborate setups to measure learning gains. At the same time though, they tend to measure learning on a very small number of topics or via a specifically-designed search tool. To what extent can we generalize the results of these small-scale studies to other domains and tasks?
5. Is it possible to measure the quality of a learning path towards the formal learning of a particular skill by taking advantage of textbooks' structures as ground truth?
6. Retrieval practice (the repeated testing of knowledge) has been shown to be beneficial to learners. Can we integrate a retrieval practice component into the search process, given that today (at least Web) search has been designed to minimize the amount of duplicate information? An added benefit of such a component: the retrieval practice questions can act as probes to test understanding and learning progress.
7. Learning through failure: users may also learn when their information needs are not satisfied and their goals are not achieved. How can we deal with that?

4.8 How should we address the conflict between the need/expectation of fast search and the need for more time in learning?

Participants: Michael Twidale, Ran Yu, Gabriele Irle, Soo Young Rieh, Heather O'Brien and Preben Hansen.

Introduction

A search engine is a quick way to get information and followed by reflection. Sometimes you also get results that require you to slow down in both the learning and searching process. It may be necessary that you look into the search result after the search. In this way the search actually continues. Therefore, the post-search activities are important. When people create something after a search activity, the search is a more engaged activity and thus, the search process is a sub-process in the learning process.In learning processes, the search activity is considered to be an instrumental activity unless the learning task is about learning how to search.

Search results. The search results should include alternative ideas that could be of interest to a person's search goal. Content-related focus in search results. Designing the "nudge" in the search activity.

Creativity. Creativity has been the focus for a long time in other disciplines such as HCI, Interaction Design and Industrial Design. These research fields have developed creativity principles, such as IDEO. Creativity and ideas may have certain characteristics, such as that ideas percolate and that creativity emerges through composition.

Creativity as part of the search process. Examples for enable randomized creative processes are, for example, IDEO's card deck and Brian Eno's (see below) creative strategies for music.

Challenges

1. Creativity and search. How we can deploy creativity (Keith Sawyer, 2013) during the search process and how to deploy creativity in our search activity.
2. To be able to design search systems so that they have different "Moods". The user could choose from these moods. How about slowing down the search system and at the same time speeding up the learning process? Design a system so that functionalities adapt to the learning process. Another dimension could be to design systems so that they show breadth and depth.
3. Playfulness in search systems. Purposeful and serendipity approaches in search systems. For example, a bookstore and a newspaper can be a metaphor for involving learning. Building a representation of a process people have been involved in.
4. How about to suggest some functionalities that perform slightly less accurate, fast, on-topic, relevance that may result in reflection.
5. Creativity in the search process. Designing for creative methods for searching as well as for search for creativity.
6. Experiment challenge: Bring two different papers from two different areas. Creative search. Predict what can emerge from merging these two.

References

1 Keith Sawyer (2013). Zig Zag: The Surprising Path to Greater Creativity. Jossey-Bass.
2 Links:
- http://www.unc.edu/home/rksawyer/
- https://global.oup.com/academic/product/explaining-creativity-9780199737574?cc=de&lang=en
- http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1118297709.html

3 IDEO: https://www.ideo.com/eu
4 Brian Eno: https://en.wikipedia.org/wiki/Oblique_Strategies

4.9 How can search systems become more conversational to include or promote learning?

Participants: Alyssa Wise, Jiyin He, Rebecca Reynolds, Dan Russell, Marcus Specht, Leif Azzopardi, Claudia Hauff.

Introduction

Commercial Web search engines are not built to support users in the type of complex searches often required in learning situations. Finding, understanding, analyzing and evaluating the documents containing information relevant to answering a complex question is a time-consuming and cognitively demanding process which requires an interactive dialogue between the search system and the user. A good conversational agent would act similar to a librarian reference interviewer who clarifies ambiguous statements, understands the context of the search and fixes category errors. Although this is currently beyond our technological abilities, in this breakout group we discussed the implications of such a conversational search approach.

Open issues

1. How should a conversational agent (presumably trained automatically on vast quantities of text) deal with questions to which no clear consensus answer exist ("Does God exist")?
2. How much of a dialogue are users willing to engage in?
3. Does the dialogue have to be explicit or can we make use of user signals that implicitly provide a dialogue response?
4. Collaborative search (several users searching together) could also be conducted with one user and several agents; how would such collaboration look like?
5. What are the functions and objectives of a conversational agent? How should a conversational agent behave when learning is the user's objective (e.g. engage the learner to not give up)? How important is the social aspect of a conversational agent?
6. Can a conversational agent act as a simulated learner to facilitate engagement with the content and help the user to clarify information? How can we design a system that incorporates deep models of learning?
7. Can we emulate how intelligent tutoring systems guide their users through the learning material? Can conversational agents guide users through those parts of the search space they have not considered before? If we do so, does that lead to more collective agreement on contentious issues in parts of the society (e.g. climate change)?

5 Panel discussions

5.1 Closing Panel Session

Claudia Hauff (TU Delft, NL), Robert Capra (University of North Carolina – Chapel Hill, US), Kevyn Collins-Thompson (University of Michigan – Ann Arbor, US), Gwen Frishkoff (University of Oregon, US), Preben Hansen (Stockholm University, SE), Noriko Kando (National Institute of Informatics – Tokyo, JP), Soo Young Rieh (University of Michigan – Ann Arbor, US), Daniel Russell (Google Inc. – Mountain View, US), and Christa Womser-Hacker (Universität Hildesheim, DE)

The closing panel discussion featured provocative questions from the participants followed by answers from the panel. We present the panel discussion as follows. First, the question raised by a person in the audience is reported, followed by different answers from the people in the panel. The panel was followed by a wrap-up session in which all participants discussed opportunities for community-building and resource sharing, proposal funding, and future collaboration.

Panel Session

The panel consisted of the following participants: Soo Young Rieh, Christa Womser-Hacker, Dan Russell, Noriko Kando, Gwen Frishkoff, Robert Capra.

Q: Where is the real impact of this research line? Is it just an academic exercise? Why now?

1. What is the impact of information on people? We have done basic Web search, we need to take the next step.
2. People have become comfortable with searching. There is a need for more sophisticated tools. Google is not designed as a learning system and yet people use it for learning. This is a high-value very specialized vertical. In this area it is really easy to try something and it fails spectacularly (e.g. Google Notebook). It is easy to build something super cool, however, it needs to be "dead simple" to be adopted. As an outsiders to the commercial search engines we need to build associative/auxiliary systems.
3. There are options for this type of research (outside of the 'Google sphere'). There are currently many 'if's' in the IS/IR research discussions. We should think about how to make them smaller.
4. We should not forget the notion of 'search as learning' as 'learning to search'!
5. Discussions focus still on the individual level; we should also consider the group level; search needs to become a societal research agenda.
6. We need a context generator ("total perspective vortex"). Social search/learning is not an echo chamber, it is a prison cell.

Q: How might we use the analogy of librarians and librarianship to be inspired for search as learning?

1. Study shows that 55% of answers by librarians are correct; but is this really the gold measure? Instead of measuring accuracy we should measure how often people come back to learn more (teaching moment)
2. Conversational agents have made it into Google, it is not a reference librarian (yet) but it can move into this direction

Q: How do you see the role of the scientific community?

1. Don't focus on things Google or Bing are good at (computations, algorithms, coding); Google isn't doing a lot of things with children search, neuro-measures, etc.

Q: What happens in the future?

1. I can ask meaningful deep questions and get synthesized complete answers.
2. Eco chamber phenomena mix learning with opinions.
3. In the future people will go back to and remise about the times without so much technology
4. Technology will be more friendly to every type of user
5. We will engage/interact with search in richer ways; richer interfaces
6. Smarter technology vs. making people get smarter; we should not get lazy
7. Ontologies are the backbone of knowledge

Position: The real essential breakthroughs will come from sensors and signal processing + AI/DNN, not better educational models or even retrieval algorithms.

1. Don't agree, we need better metadata!
2. But sensors can provide metadata!
3. Data is not the problem, making sense of it is, we need underlying models
4. Discussion on sensecam-like technologies
5. I want to search my "brain", I want to make the internal knowledge structures explicit to others e.g. in teaching

Wrap-up session

The purpose of the wrap-up session that took place as the final session was to look forward and discuss how we could take this further and how we may collaborate. This included examples of potential funding sources, additional publishing venues, datasets and evaluation frameworks, and community outreach ideas. The present Dagstuhl seminar also connects to goals and ideas described in the SWIRL report (Allen, 2012).

1. Biased structuring of search results. Think about SIGCHI, connections to interactive sense-making.
2. Some interesting resources for further connections:
   a. ASIST special interest group.
   b. PSLC DataShop learning data (https://pslcdatashop.web.cmu.edu/).
   c. ISLS: International Society of Learning Sciences (https://www.isls.org/) .
   d. CSCL (Computer-Supported Collaborative Learning).
3. A special issue on SAL for JLA (Journal of Learning Analytics) is possible.
4. Work on an idea for a Morgan & Claypool monograph.
5. A special issue on Searching as Learning was issued 2016 Journal of Information Science, Volume 42, Issue 1, February 2016
6. Funding:
   a. FET supported action: 150k to support coordination/exchange across countries, or even toward supporting a center. (NSF may have something similar reciprocal.)

b. Horizon 2020 giant funding grant scheme (January learning-related. 120 pages). 5% success rate.
c. COST: European proposal (travel funding). Innovation Training Network: partners with other countries. Existing COST KEYSTONE: Researchers from the area of database and information retrieval.
d. NSF Building Community and Capacity: data-intensive research in education, workshops.

7. University of British Columbia and Peter Wall institute offer funds for international workshops (25K).
8. There is also a need for a common dataset, evaluation framework to compare, design specific tasks, and define experimental methodology. Could make a workshop to analyze shared dataset. Make annotated data available (with inconsistencies).
9. Propose/run a TREC shared task on search as learning. For inspiration, see MMM competition: video browser showdown, or demo showdown. This could also be possible at CLEF since they has a more exploratory mode (multilingual).
10. Tutorials at SIGIR and related conferences on this area can help encourage submission of further papers in SAL.
11. Short-term exchange student research visits between universities.
12. Propose further SAL-related workshops at upcoming conferences.
13. Other communities for outreach: core education at classroom, cognitive modeling, social media, computational social science, communications, machine learning expert.

References

1 Allan, J., Croft, B., Moffat, A., Sanderson, M.: Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne. SIGIR Forum 46(1) (May 2012) pp 2–3.

Participants

- Dirk Ahlers
NTNU – Trondheim, NO
- Leif Azzopardi
University of Strathclyde – Glasgow, GB
- Robert Capra
University of North Carolina – Chapel Hill, US
- Kevyn Collins-Thompson
University of Michigan – Ann Arbor, US
- Gwen Frishkoff
University of Oregon, US
- Ujwal Gadiraju
Leibniz Universität Hannover, DE
- Tim Gollub
Bauhaus-Universität Weimar, DE
- Preben Hansen
Stockholm University, SE
- Claudia Hauff
TU Delft, NL
- Jiyin He
CWI – Amsterdam, NL
- Gabriele Irle
Universität Hildesheim, DE
- Yvonne Kammerer
IWM – Tübingen, DE
- Noriko Kando
National Institute of Informatics – Tokyo, JP
- Yiqun Liu
Tsinghua University – Beijing, CN
- Andreas Nürnberger
Universität Magdeburg, DE
- Heather O'Brien
University of British Columbia – Vancouver, CA
- Rebecca B. Reynolds
Rutgers University – New Brunswick, US
- Soo Young Rieh
University of Michigan – Ann Arbor, US
- Marc Rittberger
DIPF – Frankfurt am Main, DE
- Daniel Russell
Google Inc. – Mountain View, US
- Marcus Specht
Open University – Heerlen, NL
- Michael Twidale
University of Illinois at Urbana-Champaign, US
- Alyssa Wise
New York University, US
- Christa Womser-Hacker
Universität Hildesheim, DE
- Grace Hui Yang
Georgetown University – Washington, US
- Ran Yu
L3S Research Center – Hannover, DE