DAGSTUHL
REPORTS

Volume 9, Issue 6, June 2019

Dagstuhl Reports, Vol. 9, Issue 6 ISSN 2192-5283

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
February, 2020

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.9.6.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 19231

Empirical Evaluation of Secure Development Processes

Edited by

Adam Shostack[1], Matthew Smith[2], Sam Weber[3], and Mary Ellen Zurko[4]

1 **Seattle, US,** adam@shostack.org
2 **Universität Bonn and Fraunhofer FKIE, DE,** smith@cs.uni-bonn.de
3 **Carnegie Mellon University – Pittsburgh, US,** smweber@andrew.cmu.edu
4 **MIT Lincoln Laboratory – Lexington, US,** mez@alum.mit.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 19231 "Empirical Evaluation of Secure Development Processes". It includes a discussion of the motivation and overall seminar organization, the abstracts of the talks and a report of each working group.

Seminar June 2–7, 2019 – http://www.dagstuhl.de/19231

2012 ACM Subject Classification Security and privacy → Software security engineering, Human-centered computing, Software and its engineering → Software development process management, Software and its engineering → Designing software, Security and privacy → Systems security

Keywords and phrases security, security engineering, secure development lifecycle, empirical software engineering, empirical security engineering

Digital Object Identifier 10.4230/DagRep.9.6.1

Edited in cooperation with Daniel Votipka

1 Executive Summary

Adam Shostack
Matthew Smith
Sam Weber
Mary Ellen Zurko

The problem of how to design and build secure systems has been long-standing. For example, as early as 1978 Bisbey and Hollingworth[6] complained that there was no method of determining what an appropriate level of security for a system actually was. In the early years various design principles, architectures and methodologies were proposed: in 1972 Anderson[5] described the "reference monitor" concept, in 1974 Saltzer[7] described the "Principle of least privilege", and in 1985 the US Department of Defense issued the Trusted Computer System Evaluation Criteria[8].

Since then, although much progress has been made in software engineering, cybersecurity and industrial practices, much of the fundamental scientific foundations have not been addressed – there is little empirical data to quantify the effects that these principles, architectures and methodologies have on the resulting systems.

This situation leaves developers and industry in a rather undesirable situation. The lack of this data makes it difficult for organizations to effectively choose practices that will cost-effectively reduce security vulnerabilities in a given system and help development teams achieve their security objectives. There has been much work creating security development lifecycles, such as the Building Security In Maturity Model[1], Microsoft Security Development LifeCycle[3] OWASP[4] and ISECOM[2] and these incorporate a long series of recommended practices on requirements analysis, architectural threat analysis, and hostile code review. It is agreed that these efforts are, in fact, beneficial. However, without answers as to why they are beneficial, and how much, it is extremely difficult for organizations to rationally improve these processes, or to evaluate the cost-effectiveness of any specific technique.

The ultimate goal of this seminar was to create a community for empirical science in software engineering for secure systems. This is particularly important in this nascent of research in this domain stage since there is no venue in which researchers meet and exchange. Currently single pieces of work are published at a wide variety of venues such as IEEE S&P, IEEE EuroS&P, ACM CCS, USENIX Security, SOUPS, SIGCHI, ICSE, USEC, EuroUSEC, and many more. The idea was that bringing together all researchers working separately and creating an active exchange will greatly benefit the community.

Naturally, community-building is a long-term activity – we can initiate it at a Dagstuhl seminar, but it will require continuous activity. Our more immediate goals were to develop a manifesto for the community elucidating the need for research in this area, and to provide *actionable* and *concrete* guidance on how to overcome the obstacles that have hindered progress.

One aspect of this was information gathering on how to conduct academic research which is able to be transitioned and consumed by developers. We felt that all too frequently developer needs aren't fully understood by academics, and that developers underestimate the relevance of academic results. Our information gathering will help foster mutual understanding between these two groups and we specifically looked for ways to build bridges between them.

A second obstacle which we aimed to address is how to produce sufficiently convincing empirical research at a foundational level as well as in the specific application areas. Currently there is no consensus on what are ecologically valid studies and there are sporadic debates on the merits of the different approaches. This seminar included a direct and focused exchange of experience and facilitated the creation of much needed guidelines for researchers. In accordance with our bridge building, we also looked at what developers find convincing, and how that aligns with research requirements.

Seminar Format

Our seminar brought together thirty-three participants from industry, government and both the security and software engineering academic communities. Before the seminar started we provided participants with the opportunity to share background readings amongst themselves.

We began our seminar with level-setting and foundational talks from industrial, software engineering and security participants aimed to foster a common level of understanding of the differing perspectives of the various communities.

Following this the seminar was very dynamic: during each session we broke into break-out groups whose topics were dynamically generated by the participants. The general mandate for each group was to tackle an aspect of the general problem and be actionable and concrete: we wished to avoid vague discussions of the difficulties involved with studying secure development but instead focus on how to improve our understanding and knowledge. After each session we met again as a group and summarized each group's progress.

At the conclusion of the seminar we brought together all the participants in a general discussion about further activities. In all, a total of eighteen further activities, ranging from papers to research guideline documents, were proposed and organized by the participants.

References

1 Building security in maturity model. http://www.bsimm.com/.
2 Isecom. http://www.isecom.org.
3 Microsoft security development lifecycle. https://www.microsoft.com/en-us/securityengineering/sdl/.
4 Owasp. https://www.owasp.org.
5 Anderson, J. P. Computer Security Technology Planning Study, Volume II. Tech. Rep. ESD-TR-73-51, 1972.
6 Bisbey, R., and Hollingworth, D. Protection analysis: Final report. *Information Sciences Institute, University of Southern California: Marina Del Rey, CA, USA, Technical Report ISI/SR-78-13* (1978).
7 Saltzer, J. Protection and the control of information sharing in Multics. *Communications of the ACM 17*, 7 (1974), 388–402.
8 United States Department of Defense. Trusted computer system evaluation criteria (orange book).

2 Table of Contents

3 Overview of Talks

3.1 Experience with the Microsoft Security Development Lifecycle

Steven B. Lipner (SAFECode – Seattle, US)

Main reference Michael Howard, Steve Lipner: "The Security Development Lifecycle", Microsoft Press, 2006
URL https://blogs.msdn.microsoft.com/microsoft_press/2016/04/19/free-ebook-the-security-development-lifecycle/

Between 2002 and 2004, Microsoft made a major commitment to software security, first executing a series of "security pushes" on major products and then introducing the Security Development Lifecycle (SDL), a mandatory process for improving products' resistance to attack. After the introduction of the SDL in 2004, the process requirements were updated periodically in response to new classes of attacks and new techniques for improving product security. This presentation summarizes the philosophy and practicalities underlying the SDL and the ways it has evolved, and outlines some of the ways that the effectiveness of the SDL and similar processes can be measured.

3.2 Security in Modern Software Development

Olgierd Pieczul (IBM – Dublin, IE)

Recent shifts in software engineering make the traditional view of software security obsolete. Secure development practices rely on several aspects of software engineering such as development process, tools and languages, developer skillset and work environment which all are rapidly changing. We observe security practices lagging behind and slowing down the transformation while increasing cost and reducing overall security.

Today's software delivery cycle can be as short as days or even hours. However, the security quality and compliance processes, reviews, testing and sign offs have been designed for long waterfall-style cycles. Similarly, while applications are being developed as small, independent microservices, security practices operate at a scale of system as a whole. These tasks tend to be labor intensive, managed, executed and reviewed manually and long to perform. This results in slowing down the delivery and reducing security assurance to rudimentary, checkbox-style level only.

This presentation covers the aspects of modern software development that have major impacts on security. We also identify and describe key areas of development that need to be considered, or much more significant considered, by security research, in particular: development process, software stack, team structure, developer skills and training and code reuse.

3.3 Software Engineers are People Too: Applying Human Centered Approaches to Improve Software Development for Security

Brad A. Myers (Carnegie Mellon University – Pittsburgh, US)

Main reference Brad A. Myers, Andrew J. Ko, Thomas D. LaToza, YoungSeok Yoon: "Programmers Are Users Too: Human-Centered Methods for Improving Programming Tools", IEEE Computer, Vol. 49(7), pp. 44–52, 2016.
URL http://dx.doi.org/10.1109/MC.2016.200

Software engineers might think that human-computer interaction (HCI) is all about improving the interfaces for their target users through user studies. However, software engineers are people too, and they use a wide variety of technologies, from programming languages to search engines to integrated development environments (IDEs). And the field of HCI has developed a wide variety of human-centered methods, beyond lab user studies, which have been proven effective for answering many different kinds of questions. In this talk, I will use examples from my own and other's research relevant to security to show how HCI methods can be successfully used to improve the technologies used in the software development process. For example, "Contextual Inquiry" (CI) is a field study method that identifies actual issues encountered during work, which can guide research and development of tools that will address real problems. We have used CIs to identify nearly 100 different questions that developers report they find difficult to answer, which inspired novel tools for reverse-engineering unfamiliar code and for debugging. We used the HCI techniques of Paper Prototyping and Iterative Usability Evaluations to improve our programming tools. Through the techniques of Formal User Studies, we have validated our designs, and quantified the potential improvements. Current work is directed at improving the usability of APIs, using user-centered methods to create a more secure Blockchain programming language, addressing the needs of data analysts who do exploratory programming, helping programmers organize information found on the web, and helping end-user programmers augment what intelligent agents can do on smartphones.

References

1 Brad A. Myers, Andrew J. Ko, Thomas D. LaToza, and YoungSeok Yoon. *Programmers Are Users Too: Human-Centered Methods for Improving Programming Tools*. IEEE Computer, Special issue on UI Design, 49, issue 7, July, 2016, pp. 44-52

3.4 A Series of Experiments on Software Design Patterns

Walter Tichy (KIT – Karlsruher Institut für Technologie, DE))

Main reference Lutz Prechelt, Barbara Unger, Walter F. Tichy, Peter Brössler, Lawrence G. Votta: "A Controlled Experiment in Maintenance Comparing Design Patterns to Simpler Solutions", IEEE Trans. Software Eng., Vol. 27(12), pp. 1134–1144, 2001.
URL https://doi.org/10.1109/32.988711

Software Design Patterns are proven solutions for software design problems. They are claimed to improve software quality, programmer productivity, and communication among developers, among others. A series of three experiments tests these claims.

The first experiment checks whether the presence or absence of design pattern documentation makes a difference [1]. Subjects received programs which contained design patterns and were asked to perform maintenance tasks on them. The experiment group received a few lines (20-30) of extra documentation pointing out the design patterns. Results indicate that documenting design patterns speeds up maintenance tasks that involve those patterns, or reduces defects. This experiment has been repeated using a different programming language with consistent results.

The second experiment compares maintenance tasks on programs with and without design patterns [2]. Four programs were implemented in two versions each: One with patterns and one in a modular fashion without design patterns. The results show that not all patterns are equally effective. Some speed up pattern-relevant maintenance tasks, some have no effect, and some require extra time. The complexity of the patterns seems to play an important role. This experiment has been replicated yielding similar results.

The third experiment tests whether programmers communicate more effectively with shared knowledge of design patterns [3]. Pairs of programmers were audio and video recorded when discussing maintenance tasks. The recordings were analyzed and the contributions of each participant were counted. The results clearly show that with shared pattern knowledge, a more balanced communication results, where balanced means that team members contribute equally to the discussion.

These three experiments (and their replications) support and complement each other and confirm the hypothesized positive effects of design patterns.

(The experimental techniques employed might also be of interest. Double counterbalancing helps neutralize sequencing effects. Communication lines as a measure of effective communication might be interesting in other contexts.)

References

1 Lutz Prechelt, Barbara Unger-Lamprecht, Michael Philippsen, and Walter F. Tichy. *Two Controlled Experiments Assessing the Usefulness of Design Pattern Documentation in Program Maintenance*. IEEE Trans. On Software Engineering, 28(6), June 2002, 595-606.

2 Lutz Prechelt, Barbara Unger, Walter F. Tichy, Peter Brössler and Lawrence G. Votta. *A Controlled Experiment in Maintenance Comparing Design Patterns to Simpler Solutions*. IEEE Transactions on Software Engineering, 27(12), 1134-1144, 2001.

3 Barbara Unger, Walter F. Tichy. *Do Design Patterns Improve Communication? An Experiment with Pair Design*. May 2000, in WESS 2000. http://ps.ipd.kit.edu/176_766.php.

3.5 Empiricism in Software Engineering and Secure Systems

Laurie Williams (North Carolina State University – Raleigh, US)

The science of software engineering and the science of security can be advanced through the use of sound research methodologies while conducting empirical studies. With the use of sound methodologies, the evidence produced by a study is more credible, convincing, and substantiated and can be built upon by future researchers. As a result, the research results have more impact on other researchers and on practitioners. Additionally, meta-analysis and theory/law building is enabled when the research results are thoroughly reported.

Over the last twenty or more years, empirical software engineering researchers have conducted explicit efforts to mature the use of sound methodologies by creating guidelines and examples that bring established research practices into the context of software engineering. These guidelines include books and journal papers. Additionally, communities have come together with the goal of advancing software engineering research methods, including Dagstuhl seminars (Seminars 06262 and 10122), the International Software Engineering Research Network (ISERN), and the Empirical Software Engineering and Measurement (ESEM) conference. Over time, the top software engineering conference began to insist on the use of sound research methods which raised the bar for the community. A comparison of the research and validation methods in accepted papers in the 2002 International Conference on Software Engineering (ICSE) [1] and the 2016 ICSE demonstrated that the 2016 papers [2, 3] were more likely (3% versus 19%) to be an empirical reports which were more likely (0% versus 30%) to be accepted when compared with the 2002 papers. Additionally, the 2016 papers were more likely to have a formal evaluation (5% versus 35%) while papers that used example as their evaluation or to have no evaluation (20% and 7%, respectively, in 2002) were essentially non-existent in 2016.

Empiricism in emerging in security research. To establish a baseline, Carver et al. [4, 5] analyzed the evidence of science papers in the top security conferences, ACM Computer and Communications Security (CCS) and IEEE Security and Privacy. Their main motivation was to assess whether the papers reported information necessary for three key pillars of scientific research: replication, meta-analysis, and theory building. They examined the papers for completeness such that other researches would be enabled to understand, replicate, and build upon the results but looking for: research objectives, subject/case selection, the description of the data collection procedures, the description of the data analysis procedure, and the threats to validity. They found that 80% of the papers did not provide clearly-defined and labeled research objectives to define the goals, questions, and/or hypotheses of the research. Additionally, 70% had no discussion of the threats to validity or limitations of the work such that future researchers can make design choices to address these threats and limitations and to contextualize meta-analysis. These results indicate a need for the security research community to go through the type of maturation in empirical research that the software engineering community has gone through, including publishing guidelines in books and journal and the establishment of a community to drive this maturation.

References

1 M. Shaw, *Writing Good Software Engineering Research Papers* Proceedings of 25th International Conference on Software Engineering (ICSE'03), pp. 726-736, 2003.

2 C. Theisen, M. Dunaiski, L. Williams and W. Visser, *Writing Good Software Engineering Research Papers: Revisited* 2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C), Buenos Aires, 2017, pp. 402-402.

3 C. Theisen, M. Dunaiski, L. Williams and W. Visser, *Software Engineering Research at the International Conference on Software Engineering in 2016* SIGSOFT Softw. Eng. Notes, Vol. 42, No. 4, pp. 1-7. Buenos Aires, 2017, pp. 402-402.

4 Morgan Burcham, Mahran Al-Zyoud, Jeffrey C. Carver, Mohammed Alsaleh, Hongying Du, Fida Gilani, Jun Jiang, Akond Rahman, özgür Kafalı, Ehab Al-Shaer, and Laurie Williams. *Characterizing scientific reporting in security literature: An analysis of ACM CCS and IEEE S&P papers* In Proceedings of the Hot Topics in Science of Security: Symposium and Bootcamp, HoTSoS, pages 13–23, New York, NY, USA, 2017. ACM.

5 Jeffrey C. Carver, Morgan Burcham, Sedef Akinli Kocak, Ayse Bener, Michael Felderer, Matthias Gander, Jason King, Jouni Markkula, Markku Oivo, Clemens Sauerwein, and

Laurie Williams. *Establishing a baseline for measuring advancement in the science of security: An analysis of the 2015 IEEE Security & Privacy proceedings* In Proceedings of the Symposium and Bootcamp on the Science of Security, HotSos '16, pages 38–51, New York, NY, USA, 2016. ACM.

3.6 "Usable Security" approaches to Empiricism for Secure Software Development

Matthew Smith (Universität Bonn and Fraunhofer FKIE, DE), Sascha Fahl (Leibniz Universität Hannover, DE), and Michelle L. Mazurek (University of Maryland – College Park, US)

Usability problems are a major cause of many of today's IT-security incidents. Security systems are often too complicated, time-consuming, and error prone. For more than a decade researchers in the domain of usable security (USEC) have attempted to combat these problems by conducting interdisciplinary research focusing on the root causes of the problems and on the creation of usable security mechanisms. While major improvements have been made, to date USEC research has focused almost entirely on the non-expert end-user. However, many of the most catastrophic security incidents were not caused by end-users, but by developers. Heartbleed and Shellshock were both caused by single developers yet had global consequences. Fundamentally, every software vulnerability and misconfigured system is caused by developers making mistakes, but very little research has been done into the underlying causalities and possible mitigation strategies. In this talk we will explore the need for empiricism for secure software development in several application areas, including TLS, passwords, malware analysis and vulnerability analysis.

3.7 How the usable security community does developer studies: Field-ish studies with Build It, Break It, Fix It

Michelle L. Mazurek (University of Maryland – College Park, US)

Joint work of Andrew Ruef, Michael Hicks, James Parker, Dave Levin, Michelle L. Mazurek, Piotr Mardziel
Main reference Andrew Ruef, Michael W. Hicks, James Parker, Dave Levin, Michelle L. Mazurek, Piotr Mardziel: "Build It, Break It, Fix It: Contesting Secure Development", in Proc. of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016, pp. 690–703, ACM, 2016.
URL http://dx.doi.org/10.1145/2976749.2978382

Lab studies, field measurements, and field studies all contribute valuable knowledge to our understanding of secure development, but they also all have important drawbacks in terms of internal and external validity. The Build It, Break It, Fix It competition represents a new point in this tradeoff space. Teams compete over several weeks to build software according to a spec, gaining points for functionality and performance, then compete to break each others' software, causing the vulnerable team to lose points. The competition provides insight into how and why certain vulnerabilities arise, providing more control than a field study but more ecological validity than a smaller-scope lab study.

4 Working groups

4.1 Building Code Breakout

Adam Shostack and Carl Landwher

Joint work of Adam Shostack, Carl Landwehr, Florian Alt, Michael Coblenz, Serge Egelman, Shamal Faily, Joseph Hallett, Trent Jaeger, Mike Lake, Steve Lipner, Fabio Massacci, Simon Ou, Olgierd Pieczul, Riccardo Scandariato, Laurens Sion, Laurie Williams, Sam Weber

Building codes for physical structures have been developed over centuries to address societies' needs for buildings that are safe for their occupants and can stand up to customary natural threats – rain, wind, fire, earthquakes – for the areas in which they are built. Although such codes are diverse, they generally call for an initial approval by a local authority of the design and specification of the proposed structure, use of approved materials and methods in its construction, inspections to be carried out by trained third-party inspectors during construction, and a final approval for occupancy before the building can be occupied.

The construction and deployment of software systems, with few exceptions, is not subject to these kinds of third-party inspections and approvals. It has been proposed that the kinds of mechanisms found in the building code model might help achieve better security in the software systems deployed in, for example, medical devices, power grids, and "Internet-of-Things" devices.

The group considered what limits the building code metaphor might have in relation to software systems, what hypotheses about this approach might be developed and subjected to empirical evaluation, and what a roadmap might look like that would lead to the appropriate development and use of building codes for software systems with security and safety responsibilities.

4.1.1 Limits to the Metaphor

The discussion revealed many ways in which building codes for physical structures might reasonably be applied to software structures with security and safety responsibilities. The blueprint for a building specifies a structural design and plan for implementation and must be approved by a local authority before construction can proceed. Documents capturing system and software requirements and design have a similar role in software construction, though often implementation may proceed based on knowingly incomplete specifications and designs. Changes in the software requirements and design as construction proceeds are common in software, but they also occur in building construction.

Building codes normally require sequenced inspections of aspects of the building during construction: footings and foundations must be inspected before floors and walls are raised; plumbing and heating systems must be inspected before the walls are closed up and this work is hidden, and so on. Finally, when construction is complete a final inspection is required before the local authority certifies that the building can be occupied. As software systems are built there are often code reviews, unit and subsystem tests, and finally system testing before the software is released. In some environments, an authority may required to issue a final approval to operate before the system is made available to its users.

The metaphor was found to be limited in that software has no physical manifestation beyond the bits representing the programs, software can be replicated nearly without cost, and it can be transported easily around the globe. Buildings of course have substantial mass and are not easy to alter, move, or replicate. Physical structures change, but much more slowly than software, much of which is subject to continual updates. Inspectors of physical structures require some training typically measured in weeks or months; inspectors of software generally require years of training.

The group's conclusion was that despite the limitations, the building code metaphor seems to apply well to software with security responsibilities.

4.1.2 Hypotheses to Test / Questions to be Answered

The discussion on generating hypotheses to test opened by considering some characteristics desired of a proposed building code. As a principle, the building code should result in secure code and not in evidence to be consumed by the code inspector. Whatever evidence has to be tied to the code and the close the better. It was observed that the Common Criteria scheme has led to development of required documentation after the fact. These documents are then reviewed without reference to the actual underlying software, resulting in relatively low assurance of security at relatively high cost. Future building codes for software should avoid this trap.

The Windows Logo program, which enables developers of Windows applications to display the Windows logo only if their software passes certain automated tests, and the Apple Store program, which only accepts applications that pass both automated tests and some human review, were discussed and found to be within the category of software building codes.

In the context of an imagined building code for software with security responsibilities:

1. Would an inspector be required to exercise judgement, or could inspection be effectively automated?
2. What type of specification would be sufficient as the basis for effective enforcement (e.g., would a set of use cases or "user stories" be adequate? Would a data flow diagram specification suffice?
3. Does increasing the number of acceptance criteria incentivize better designs?
4. Will implementing an SDL reduce errors in implementations?
5. Will having a developing and applying a building code be better than doing nothing at all?

4.1.3 Working Notes and Ideas to Develop

1. Examples – Are the Apple store and Windows Logo "building codes"? What about the UK's DMCS IoT standard?
2. Principle: The building code should result in secure code and not in evidence to be consumed by the code inspector. Whatever evidence has to be tied to the code and activities intended to deliver products, and the closer the better. (Contrast: work done solely to satisfy the building inspector.)
3. Experiments on the building code: at every iteration a product developed to meet the code should be distinguishable from a product that has not been produced accordingly. Eventually we should be able to demonstrate that software systems built to the code were not just distinguishable, but better than without a building code.
4. Introduction of new technology for which verification is not applicable could be tested in unregulated area first and then moved to regulated areas when tools becomes available. Also, what happens when "the laws of physics" change?
5. Scalability of "testin" wrt the cost of the testing (e.g. CE label). Do we know if the types of "testing" experiments that we can run are linear or with exponentially decreasing returns? So that if you achieve a threshold that would enough because the next level

would cost you exponentially more (in either time or costs) while only delivering minor increasing quality.

6. Domain and time applicability Should a code commit to a building code version at production time and stick to it or must comply with the latest version at delivery time? What about "capricious" rules? What about late-breaking changes to the laws of physics? Building codes don't need to deal with those.
   We should distinguish between the "experiments" on the final product and "experiments" on the production process

4.1.4 Building Code Components

Building codes vary widely around the world, but in this subsection we describe a typical code for physical structures, which has aspects that occur at multiple phases of construction.

Before a building is even planned, separate goals are defined for structural integrity, fire safety, plumbing, drainage, electrical, HVAC and ADA compliance. Building codes constrain both a new building's specification and its design. After a building is designed, permitting is done with the municipality with paper copies of the documents being filed and sometimes a physical examination of the site is done. At each phase of physical construction (foundation laying, framing, etc) on-site inspections are done. When completed, a certificate of occupancy needs to be obtained, which also involves a separate inspection. Additionally, in order to maintain compliance with the building code, additional inspections are required over time to ensure that elevators, major appliances and such-like continue to be safe.

4.2 Ecological validity and study design for empirical secure development studies

Michelle L. Mazurek (University of Maryland – College Park, US) and Daniel Votipka (University of Maryland – College Park, US)

Joint work of Michelle L. Mazurek, Daniel Votipka, Tamara Denning, Heather Richter Lipford, Adam Aviv, Tobias Fiebig, Mathew Smith, Brendan Murphy, Michael Coblenz, Eric Bodden, Sascha Fahl, Mary Ellen Zurko, Walter F. Tichy, Walter, Brad Myers, Charles Weir

The design of a study is critical to producing correct results and the secure development researcher faces several unique and difficult challenges when attempting to choose the right design. For example, most development tasks are performed over weeks of collaborative work, making it difficult to even simulate ecologically valid conditions in a controlled lab setting. Also, security is typically a secondary focus in real-life settings, requiring further complication of the study design to avoid unrealistic behaviors. Further, measuring *correct* behaviors with respect to security is difficult as it generally requires proving the *absence* of any errors. Developer experience can also vary widely across many dimensions including programming ability, development workflow and security knowledge, reducing the generalizability of results without a broad sample of participants.

Fields such as software engineering [4] and usable security [5] offer recommendations for study design to resolve some of these issues. Other recent work has focused on specific challenges of secure development measurement, i.e., whether remote participants produce similar results as local participants; on the differences between using students, GitHub users, freelancers, and corporate developers [1, 2, 3]; and the effect of varying incentive structures [3].

Unfortunately, little general guidance specific to secure development study design currently exists for newer researchers, and several specific questions still remain open. Therefore, there is a need both for outlining best practices for study design based on work in other fields, and also investigating open questions specific to secure development tasks. As a first step toward addressing these issues, this working group focused on three specific tasks:

- Define a set of generic guidelines for new researchers and highlight security-specific examples of their application.
- Identify methodological issues requiring further research
- Identify a set of possible metrics for measuring aspects of secure development

4.2.1 Design Guidelines

The first group identified an initial set of generic guidelines useful to empirical research broadly, along with secure-development-specific examples for each. We list these guidelines and examples here.

1. Leverage well-established study-design practices from other fields

   Many of the decisions faced when designing a empirical evaluation of secure development are not unique to this domain of research. Other fields such as software engineering, human-computer interaction, psychology, sociology, and anthropology have established best practices that address many of these issues. For example, the *Guide to Advanced Empirical Software Engineering* offers best practices for topics such as survey design, statistical analysis, and research ethics. Researchers should consider these best practices for decisions that are not specific to secure development research.

2. Use pilot studies to hone in on the research objective and be willing to iterate on the study design

   The initial study design is rarely perfect. The research team typically does not have the necessary domain knowledge to predict how participants will interpret survey questions, task descriptions or interface element or how they will respond during the study. Therefore, it's important that all studies begin with a pilot where participants can share their impressions of the study to ensure the design actually meets the stated research objectives or whether new objectives should be sought. Similarly, as the study is being carried out, the researchers should monitor data collected and be open to making updates to the design if it becomes evident that changes are necessary.

3. Consider the metrics appropriate for the study methodology and desired outcomes

   One of the most important considerations when designing any study is in determining what data to collect. This includes what data to collect about the participant themselves and study-specific data, e.g., interactions with a tool or survey responses. Significant care should be taken to ensure that both the variables under study along with any potentially confounding effects are measured as precisely as possible. This is particularly difficult in security where *correct* development solutions require the absence of errors. Therefore, the researcher must consider designs that limit the scope of possible *correct* solutions, making analysis more manageable.

4. Consider the level of expertise of participants

   Because developers can vary drastically with respect to their secure-development expertise, it is important that the research team consider whether and how these variations will affect the study. For example, when A/B testing a tool to support vulnerability discovery, the researcher should consider the participants' security expertise. Because developers with more expertise are more likely to find a vulnerability irrespective of the tool being tested, their expertise should be measured and considered as a covariate during analysis. Alternatively, the participants could be provided with additional training at the start of the experiment to ensure all participants begin on equal footing.

5. Create a context for the experiment that appropriately prompts participants

   Participants' behaviors are typically context-dependent. This is especially true for security and privacy decisions [6]. Thus, the context in which participants are surveyed or observed should be selected to best fit the experiment's goals. For example, a researcher whose goal is to measure the impact of a new penetration testing tool at improving vulnerability discovery could run the study as part of a Capture-the-Flag exercise, where participants are instructed to exploit a series of programs, encouraging participants to think like an attacker.

6. Consider how and when to use deception to hide the true purpose of your study

   In some cases, informing participants of the true nature of a study could bias them to behave differently than they would in a similar real-life scenario. For example, participants may be unwilling or ashamed to admit they do not consider security. However, the use of deception significantly complicates the study design and in general, the research should strive to be as transparent with participants a possible (as discussed in the next guideline). Thus, researchers should balance these needs by considering when deception is necessary to maintain the study's integrity.

7. Consider ethics in the design of the study

   It is morally imperative that experiments on human subjects be performed ethically. Therefore, it is important that the researcher consider potential ethical dilemmas during the study's design. For example, studies of habituation to compiler security warnings could cause developers to be more likely to ignore these types of problems in their normal work. Researchers should consider these possible side effects and design mitigations against them. The principles of the Belmont Report–*Respect for Persons*, *Beneficence*, and *Justice*–and the Menlo Report, which builds on the Belmont report focusing on information and communications technologies, provide useful guidelines for ethical considerations.

4.2.1.1 Future work

Future work will look at refining this list and providing further specific design examples for secure development studies. We will look at collecting these guidelines both from seminar participants and the broader community into an online living document. This document could provide a useful reference point for researchers new to the field.

Table 1 Open questions for secure software development study methodology.

Question	Amount Currently Known
How much specification or documentation should be provided for each task? Should unit tests be provided?	○
How much prompting for the importance of security should be provided?	◐
How much time should be provided to complete the task? A fixed time budget or unlimited time?	○
Task complexity and fidelity to real-world tasks:	
Reading code vs. writing code	○
Real-world code vs. contrived experimental code	○
From-scratch development vs. editing pre-written code	○
Should participants receive feedback and then get a second chance to improve their code?	◐

4.2.2 Studying Methodology

The second group considered open questions with regard to study design for secure development. The goal was to brainstorm empirical studies that could shed light on how to make appropriate study-design choices in secure-development studies.

Table 1 reviews the open questions the group considered. For each question, the group characterized what is currently known about this question: a significant amount (●), some (◐), or little to none (○).

Further discussion on the general idea of how to think about task instructions and fidelity suggested several potential concrete studies:

- What does real-world tasking in industry look like (interview study)
- Comparing writing secure code to fixing insecure code
- Adding irrelevant details to the task specification to increase cognitive load
- Comparing different amounts of security framing and/or role-playing, to similar tasks without context
- Comparing convincing deception, unconvincing deception, and no deception
- Comparing starting from scratch to adding new features, with different size additions
- Comparing results when task code is written on paper, with a text editor, with an IDE, and with a debugger
- Comparing requiring participants to use a standard development setup to allowing them to use their own preferred setup
- Repeating any of the above studies across a variety of security contexts (e.g., not just cryptography)

4.2.3 Metrics for Studying Secure Development

The final group explored features of developers and the organizations they belong to that might be useful in future work. Such features could be measured as outcome variables, or used as covariates in study design. For each identified feature, the group sought to identify existing metrics used to measure these features. This brainstorming exercise highlighted several metrics that could be adopted by secure development researchers. For example, this included measurements for general security awareness (Human Aspects of Information Security Questionnaire (HAIS-Q) [11]), security behavior (Security Behavior Intention Scale (SeBIS) [7]), and developer personality (Five-Factor Personality Inventory [12], Consideration for Future Consequences (CFC) [10], Need for Cognition (NFC) [9], Domain-Specific Risk-Taking scale (DoSpeRT) [8]). The group also suggested further investigation of psychology, sociology, and anthropology literature to identify metrics for features such as emotional state, frustration towards security, and organizational culture. Finally, the group identified some features that have historically only been measured using simple, Likert-scale self-report questions. This included expertise in development and security. Some work has attempted to use knowledge assessments to gauge these expertise levels, but these tests tend to be cumbersome, and there is no clear evidence of their validity for measuring expertise beyond the specific questions asked. Future work should consider more efficient alternatives.

References

1 Yasemin Acar, Christian Stransky, Dominik Wermke, Michelle L. Mazurek, and Sascha Fahl. Security Developer Studies with GitHub Users: Exploring a Convenience Sample. In *Proceedings of the 13th Symposium on Usable Privacy and Security (SOUPS)*, pages 81—95, 2017.

2 Alena Naiakshina, Anastasia Danilova, Christian Tiefenau, and Matthew Smith Deception Task Design in Developer Password Studies: Exploring a Student Sample. In *Proceedings of the 14th Symposium on Usable Privacy and Security (SOUPS)*, pages 297—313, 2018.

3 Alena Naiakshina, Anastasia Danilova, Eva, Gerlitz, Emanuel von Zezschwitz, and Matthew Smith. "If You Want, I Can Store the Encrypted Password": A Password-Storage Field Study with Freelance Developers. In *Proceedings of the 37th Conference on Human Factors in Computing Systems (CHI)*, pages 140—152, 2019.

4 Forrest Shull, Janice Singer, and Dag I.K. Sjøberg. Guide to Advanced Empirical Software Engineering. *Springer-Verlag*, 2007.

5 Stuart Schechter. Common Pitfalls in Writing about Security and Privacy Human Subjects Experiments, and How to Avoid Them. *https://cups.cs.cmu.edu/soups/2010/howtosoups.pdf*, 2010.

6 Helen Nissenbaum. Privacy as Contextual Integrity. *Washington Law Review*, pages 119–157, 2004.

7 Serge Egelman and Eyal Peer. Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS). In *Proceedings of the 33rd Conference on Human Factors in Computing Systems (CHI)*, pages 2873–2882, 2015.

8 Ann-Renee Blais and Elke U. Weber. A Domain-Specific Risk-Taking (DoSpeRT) Scale for Adult Populations. *Judgement and Decision Making*, vol. 1, no. 1, pages 33–47, 2006.

9 John T. Cacioppo, Richard E. Petty, and Chuan Feng Kao. The Efficient Assessment of Need for Cognition. *Journal of Personality Assessment*, vol. 48, no. 3, pages 306–307, 2010.

10 Jeff Joireman, Monte J. Shaffer, Daniel Ballet, and Alan Strathman. Promotion Orientation Explains Why Future-Oriented People Exercise and Eat Healthy: Evidence From the Two-

Factor Consideration of Future Consequences-14 Scale. *Personality and Social Psychology Bulletin*, vol. 38, no. 10, pages 1272–1287, 2012.
11 Kathryn Parsons, Dragana Calic, Malcolm Pattinson, Marcus Butavicius, Agata McCormac, and Tara Zwaans. The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies. *Computer & Security*, vol. 66, pages 40–51, 2017.
12 Paul Costa and Robert R. McCrae. Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional Manual. *Psychological Assessment Resources*, 1992.

4.3 Clever Recruitment Techniques: How to Design Studies that get Enough of the Right Kind of Participants

Tobias Fiebig (TU Delft, NL), Michael Coblenz (Carnegie Mellon University – Pittsburgh, US), and Fabio Massacci (University of Trento, IT)

Joint work of Luigi Lo Iacono, Sascha Fahl, Michael Coblenz, Tobias Fiebig, Adam J. Aviv, Florian Alt, Tamara Denning, Daniel Votipka, Michelle Mazurek, Fabio Massacci

4.3.1 Introduction

Recruiting participants for studies among professionals is a serious obstacle for studies on programmers' security behavior and practices. In the past, researchers from the usable security community tried to use mass mailing campaigns to app developers [1], community driven approaches [2], and recruiting among students [3]. In this workshop, we investigated recruitment techniques and incentives for professionals, and found that they are tied to the underlying research questions and design of studies on developers' security behavior.

4.3.2 Key-Insights and Research Objectives

The key-insights of our workshop are:

- The right way to recruit depends on the research question of a study, which determines the population we want to draw from, and thereby the channel and incentives we can use for recruitment.
- Our knowledge of developer types (as well as other involved stakeholders) is limited. Hence, future qualitative study should work towards a taxonomy of stakeholders and programmer types in software development.
- Our knowledge of existing recruitment channels and incentive models is mostly incidental. As such, a structured literature study should enable a more informed view on what worked in the past.
- With the currently unclear situation, extensive pre-testing is essential for finding the right combination of recruitment methods and target population. a future study should aim at validating selected combinations of methods and target groups for common research question types.

4.3.3 Common Research Question Structures

First, we explored common research question setups, to explore the context of studies on secure software development. We identified four distinct research question setups:

1. How does a certain artifact/organizational structure/training X promote or not promote better secure outcomes Y?
2. What are the incentives/barriers X to improving security in a particular operational condition Y?
3. What are the challenges/problems faced by population X, and how can they be identified/mitigated/denied?

We make two observations based on these general research questions: (i) While we selected general descriptions, in practice instanciations of these question may have a substantial conceptual overlap, and, (ii) The setup of the research question may already determine the population we have to select our sample from.

4.3.4 Population Types and Roles

Next, we attempted to categories the types and roles involved in producing software.

- Managers
- Programmers as users
- Testers
- Security teams
- Designers
- Domain experts
- Tool creators
- Users
- Programmers as creators
- Architects/Requirements engineer
- Code reviewers
- Hackers
- Attackers

We note that we do not have enough validated insights into the roles commonly involved in software development. The above list is tentatively based on an in-group brainstorming. While, technically, these types and roles can be derived from descriptive literature of the software development process, this does not necessarily reflect how these are set-up in practice. Hence, we suggest further qualitative studies, combined with literature research, to identify types and roles, and validate them in practice.

4.3.5 Incentives

Next, we discussed incentives for study participation. Even though our research could be better informed by, e.g., Motivational Theory from psychology and social science, we note that the field of software development usually changes frequently. Hence, insights from other fields, especially if they are older, may not transfer directly. Incentive methods we discussed are:

- **Cash:** Either as a flat-payment, a bonus for completion, a raffle, per-hour, as a prize, or as additional staff time (if mandated by leadership). A bonus for good performance might mitigate boredom during the study.
- **Fame**, i.e., by combining the survey with a leader board. Competing with others in general (Vanity/Victory).
- Enabling a **rare experiences** and/or Fun for participants, e.g., driving in a race car, joining a key-note, etc.

- The prospect of **helping others**.
- Helping ones' own organization / Oneself.
- Building better (own) system.
- An opportunity for networking, **social gains**, and competence gains, i.e., by obtaining new technical skills or exposure to potential interesting technology.
- **Distraction** from ones work.
- Obtaining **Credits** (students).
- Learning **best practices** for (own) studies.

We note that some of these incentives may indeed be rather expensive. Others, especially unique experiences, may be more readily available to academic researchers than expected. For example, several universities have student teams operating FormulaE or Forzza race cars.

4.3.6 Recruitment Channels

Finally, we discussed recruitment channels.

- *Community engagement:* For example, in-person participation at hacking events, present at industry conferences, using twitter/slack with hashtag #organization, retweeted by the organization. These are difficult to scale and not necessarily sustainable.
- *Going to professional events focusing on recruitment:* This can be supported by snowballing at conference (with quick ways, recording the conversation), by having a booth, or by giving a talk at conference, asking the audience to participate. Furthermore, it might be possible to integrate the survey in the talk, e.g., by using mentimeter (mentimeter.com) with live display during the presentation.
- *Top-down contact within industry/professional organizations:* Letting industry organizations contact their members might be more incentivizing to professionals. This is especially relevant for questionnaires that should be answered by managers. In general, involving management first seems to provide reasonable results.
- *Class projects with optional participation to the study:* This primarily leverages ongoing classes, but one might also use mailing lists for previous courses and advertising through other classes. A participant notes that lots of people give interest but do not follow up on this.
- *Recruit students locally:* For example with tear-off strips, and strategically placed, e.g., in start-up hubs, or via local mailing list (academic organization) and flyers.
- *Crawling public information:* For example, crawling and contacting Github users and Google play (App authors).
- *Hiring:* Via platforms like Mechanical turk and other freelancer platforms.
- *Building a panel of people interested for future studies:* This can be combined by combining going to conference, and asking participants if they are interested in future studies, and getting a bunch of people to keep up with the panel.
- *Engaging Local companies:* For example, by integrating them in thesis/internship programs, and directly recruiting CTOs/CIOs for interviews.
- *Release an application and wait until people actually use it.*
- *Offering Trainings:* This uses "hands-on" as an exercise, building on a within subject design.
- *Go where people are waiting:* If a significant wait time is involved, people might seek some distraction. The entrance queues at developer conferences might be handy here.
- *Facebook Ads:* These adds usually cost a few hundred dollars, and the standard answer rate is around 3%; Completion of the survey may be optional.

- *Send personal invitations:* Mass mails have a specific “smell” to them, which might disincentivize participation. Making them less standardized might induce a sense of personal engagement, leading to higher conversion.

4.3.7 Summary and Further Work

In summary, we find that pre-test and continuous evaluations are necessary for each experimental design. This should be an iterative process with several pre-tests. For further work, we suggest to test what incentives work with which populations for which recruitment techniques. In general, this will not be possible for an exhaustive list. Hence, this study should be informed by the most common research questions and their relevant populations. The study should build on motivation theory from psychology and social science. Similarly, we have to build a taxonomy of populations, i.e., types of developers and other involved stakeholders. This may be further informed by systematizing knowledge on which combinations of populations, recruitment methods and incentives prove successful in the literature.

References

1 Yasemin Acar, Michael Backes, Sascha Fahl, Doowon Kim, Michelle L Mazurek, and Christian Stransky. You get where you’re looking for: The impact of information sources on code security. In *Proceedings of the IEEE Symposium on Security and Privacy (Oakland)*, pages 289–305, 2016.
2 Constanze Dietrich, Katharina Krombholz, Kevin Borgolte, and Tobias Fiebig. Investigating System Operators’ Perspective on Security Misconfigurations. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 1272–1289, 2018.
3 Katharina Krombholz, Wilfried Mayer, Martin Schmiedecker, and Edgar Weippl. “I Have No Idea What I’m Doing”-On the Usability of Deploying HTTPS. In *Proceedings of the USENIX Security Symposium (UseSec)*, pages 1339–1356, 2017.

4.4 API Usability Heuristics

Matthew Smith (Universität Bonn and Fraunhofer FKIE, DE) and Joseph Hallett (University of Bristol, GB)

Security API heuristics can give some metrics about what security problems reside in code. The term has slightly different meanings to different communities however, but generally security *principles* or *heuristics* are smaller, more general rules independent of technology, whereas security *guidelines* are larger and more specific, and security *standards* have been agreed upon by a larger community. It is also not entirely clear what a *security API* is, as the term can include not just cryptographic and security related code but also APIs where the security aspects are less obvious; for instance REST APIs and file access dialogues, and data structures.

There are many different problems we might hope a security API heuristic might be able to detect. Problems include issues with *mental models* or *naming conventions* where a developer’s assumptions lead them to believe that APIs disallow (or allow) certain behaviour, such as the ability to access local files from a URL interface; issues with *adversarially controlled input* where a developer mistakenly believes that an input will always be sanitised.

Issues with *documentation* are common where documentation is missing or wrong; as well as issues with canonicalization, redundancy, immutability and call order. Even if an API is correct at one time, updates to library APIs can lead to functions and patterns becoming deprecated and needing updates, and building and linking a library is something many developers struggle with.

Security heuristics, such as Green and Smith's rules for a good crypto API suggest principles to help fix these problems but how might we evaluate their effectiveness? A/B testing is hard to use as the sheer number of parameters can lead to unreasonably large numbers of participants and tests being performed. Instead other approaches such as qualitative evaluation of heuristics against APIs, checking known API problems against proposed heuristics, and analysing change logs to see how APIs were adapted in response to a usability problem may provide better results.

Future work will look at collecting API heuristics and guidelines into an online resource, that would also collect new proposals for heuristics and the results of studies using the heuristics. Other work will look at running experiments to see whether students produce more secure APIs when given heuristics than without them. Finally if a study shows that a significant number of people fail to use an API correctly because of a usability issue, then we ought to be able to report and track this issue–perhaps a new *usability* CVE database might be a sensible way forward.

4.5 Software Security Metrics

Eric Bodden (Universität Paderborn, DE)

Joint work of Eric Bodden, Michael Coblenz, Tamara Denning, Shamal Faily, Joseph Hallett, Mike Lake, Riccardo Scandariato, Reinhard Schwarz, Adam Shostack, Laurens Sion, Charles Weir, Laurie Williams

How do we know that a system is really secure? Currently, a common notion seems to be that a software is secure if it is free of vulnerabilities. But given that definition, it is now clear to practitioners that no software is secure, as *some* vulnerabilities will always be present – be it on the architectural level, in the coding or through the execution environment. In this working group we therefore discussed that metrics for software security should take into account that one should *assume breach*, i.e., a secure system must be able to withstand attacks *despite* known or unknown vulnerabilities. This is somewhat akin to the established notion of resilience, which in the past has been mostly associated with safety, not security properties. Security, if defined that way, is not a binary property but instead is a value on a gradual scale: a system can be more secure than another, but it probably cannot be secure per se. While this might sound disappointing, it very much reflects reality.

There are different stakeholders with respect to which such metrics could be established. Software architects and decision makers in software engineering would probably appreciate structural metrics over software artifacts that correlate well with security. Such metrics could include the fraction of *security code*, i.e., code that implements security features, of the overall application code. Another metric would be how many issues (of a given category) were found in a (potentially tool-assisted) code inspection of some given subsystems. This metric could then be weighted depending on the nature of this subsystem.

Some other–potentially more promising–metrics can be defined from an attacker perspective. With respect to the assume-breach paradigm, an interesting metric is "How many vulnerabilities in combination does an attacker need to exploit to fully own the system (and how hard are these to exploit)?" For many current system, this number will be one! In fact, participants in the breakout group hypothesized that every software system has an "Achilles heel" in the sense that there is a single line of code in the system that if disabled or modified all security is off. If this is true then secure systems must do everything in their power to protect such subsystems. Other attack-centric metrics include measuring an "attack surface" by tracking taint flows from inputs to security-critical components. With respect to long-term security, it is interesting to measure how easily a potentially or likely vulnerable component (say, a complex parser) could be hardened or replaced.

With these metrics as with all others it is currently an interesting and open research questions of how well they correlate with a sensible notion of security. It thus seems that the community would need to establish...

1. an agreed-upon, well-defined and measurable notion of security
2. a set of well-defined metrics with clear descriptions that metrics yield reproducible results
3. a number of empirical studies that seek to correlate these metrics (2) with the notion of security from (1)

Those metrics that correlate best could then yield effective software security metrics.

4.6 Methods For Empirical Studies of SDLs

Sam Weber (Carnegie Mellon University – Pittsburgh, US), Adam J. Aviv (U.S. Naval Academy – Annapolis, US), Michael Coblenz (Carnegie Mellon University – Pittsburgh, US), Tamara Denning (University of Utah – Salt Lake City, US), Shamal Faily (Bournemouth University, UK), Mike Lake (CISCO Systems – Research Triangle Park, US), Steven B. Lipner (SAFECode – Seattle, US), Michelle L. Mazurek (University of Maryland – College Park, US), Xinming (Simon) Ou (University of South Florida – Tampa, US), Olgierd Pieczul (IBM Research – Dublin, IE), Charles Weir (Lancaster University, UK)

In this breakout group we discussed appropriate methodologies for empirically studying secure development processes. It would be ideal if we, as a community, could produce a guide to researchers on what techniques and methodologies to use in order to empirically evaluate competing processes for secure development.

Unfortunately, the current state of the art falls far short of where we need to be in order to produce such a guide. Even such apparently easy process changes, like removing a static analysis from the acceptance tests, turn out in practice to be surprisingly subjective and highly context-dependent. Other changes are harder to evaluate: if we see more defects, are we creating more or are we better at detecting them? It was clear to the group that effective research methodologies was itself a non-trivial research topic.

In order to make a more concrete process, we discussed both hypotheses about secure development that the community maintains, but would like more concrete evidence for, and what the important outstanding research questions are. We then did a deep-dive on one of them (threat-modeling).

Things that the community believes are true, but which could use empirical evidence, include:

- threat-modeling is very important,
- vulnerabilities and vulnerability mitigation tasks are distinct from each other and need to be tracked as such in the bug tracking system, and
- securing third-party code/libraries is critically important.

Outstanding research questions (many of which are related to the above) include:

- the costs and benefits of security-related choices like the use of type-safe languages,
- the impact of various software architectures on security,
- what are the key advantages of threat modeling? Finding real threats, supporting communication, focusing thinking, allowing non-security-experts to have input into the process?
- When deciding whether or not to incorporate third-party code, what due-diligence processes are effective?
- how do processes like penetration testing affect company security culture, and how to distinguish the effects of community culture from specific processes?

In our threat-modeling deep-dive, we realized that there were a myriad of desirable properties of a threat modeling methodology, such as being low-cost, or easy to use by non-security experts. There are also many possible desirable properties of a threat model, such as identifying only feasible attacks, or enabling estimating of the cost of exploitation. A number of experimental methods for evaluating different threat modeling methodologies were brought up, including

- Obtaining a "gold standard" for a given scenario and comparing the results produced by each method,
- Ethnographic methods, where real-world teams are observed,
- Leverage the training process of companies that are bringing in external threat-modeling trainers and
- Conducting internal analyses, where an organization collects threat-modeling artifacts and then later reviews them versus vulnerabilities that were found or missed. As the actual threat model is likely to be sensitive, it probably cannot be released, but the comparison should be able to be disclosed.

The group agreed that none of these methods were "better" than the others, but rather that all of them obtain different views of threat modeling. By using disparate techniques a more complete picture can be formed. Although time prevented us from doing similarly deep dives on other problems, the group felt that this was probably a general observation: given the nature of cybersecurity, using a variety of techniques will be almost certainly required to give a complete picture of the benefits and drawbacks of any particular development methodology.

4.7 Publication or reviewing guidelines; Establishing a baseline for evidence of science in security

Laurie Williams (North Carolina State University – Raleigh, US)

To build a science, researchers need to document scientific evidence of their work so that future researchers can support or evolve the hypotheses and theories documented in published reports and papers. Carver et al. analyzed papers that appeared in two top security conferences, ACM Computer and Communications Security (CCS) and IEEE Security and Privacy (S&P) [1, 2]. Their main motivation was to assess whether the papers reported information necessary for three key pillars of scientific research: replication, meta-analysis, and theory building. To perform this work, Carver's team utilized a rubric with which they analyzed the papers: http://carver.cs.ua.edu/Studies/SecurityReview/Rubric.html. The group discussion revolved around looking at the components of this rubric and discussed the benefits of providing rubrics to enable researchers to publish papers that contain the scientific evidence others can build upon.

References

1 Morgan Burcham, Mahran Al-Zyoud, Jeffrey C. Carver, Mohammed Alsaleh, Hongying Du, Fida Gi- lani, Jun Jiang, Akond Rahman, özgür Kafalı, Ehab Al-Shaer, and Laurie Williams. *Characterizing scientific reporting in security literature: An analysis of ACM CCS and IEEE S&P papers* In Proceedings of the Hot Topics in Science of Security: Symposium and Bootcamp, HoTSoS, pages 13–23, New York, NY, USA, 2017. ACM.

2 Jeffrey C. Carver, Morgan Burcham, Sedef Akinli Kocak, Ayse Bener, Michael Felderer, Matthias Gander, Jason King, Jouni Markkula, Markku Oivo, Clemens Sauerwein, and Laurie Williams. *Establishing a baseline for measuring advancement in the science of security: An analysis of the 2015 IEEE Security & Privacy proceedings* In Proceedings of the Symposium and Bootcamp on the Science of Security, HotSos '16, pages 38–51, New York, NY, USA, 2016. ACM.

Participants

- Florian Alt
Universität der Bundeswehr – München, DE
- Adam J. Aviv
U.S. Naval Academy – Annapolis, US
- Eric Bodden
Universität Paderborn, DE
- Michael Coblenz
Carnegie Mellon University – Pittsburgh, US
- Tamara Denning
University of Utah – Salt Lake City, US
- Serge Egelman
ICSI – Berkeley, US
- Sascha Fahl
Leibniz Universität Hannover, DE
- Shamal Faily
Bournemouth University, GB
- Tobias Fiebig
TU Delft, NL
- Joseph Hallett
University of Bristol, GB
- Trent Jaeger
Pennsylvania State University – University Park, US
- Mike Lake
CISCO Systems – Research Triangle Park, US
- Carl E. Landwehr
George Washington University – Washington, US
- Steven B. Lipner
SAFECode – Seattle, US
- Luigi Lo Iacono
FH Köln, DE
- Fabio Massacci
University of Trento, IT
- Michelle Mazurek
University of Maryland – College Park, US
- Brendan Murphy
Microsoft Research – Cambridge, GB
- Brad A. Myers
Carnegie Mellon University – Pittsburgh, US
- Xinming (Simon) Ou
University of South Florida – Tampa, US
- Olgierd Pieczul
IBM Research – Dublin, IE
- Heather Richter Lipford
University of North Carolina – Charlotte, US
- Riccardo Scandariato
Chalmers and University of Gothenburg, SE
- Reinhard Schwarz
Fraunhofer IESE – Kaiserslautern, DE
- Adam Shostack
Seattle, US
- Laurens Sion
KU Leuven, BE
- Matthew Smith
Universität Bonn and Fraunhofer FKIE, DE
- Walter F. Tichy
KIT – Karlsruher Institut für Technologie, DE
- Daniel Votipka
University of Maryland – College Park, US
- Sam Weber
Carnegie Mellon University – Pittsburgh, US
- Charles Weir
Lancaster University, GB
- Laurie Williams
North Carolina State University – Raleigh, US
- Mary Ellen Zurko
MIT Lincoln Laboratory – Lexington, US

Report from Dagstuhl Seminar 19232

Ubiquitous Computing Education: Why, What, and How

Edited by

Audrey Girouard[1], Andrew L. Kun[2], Anne Roudaut[3], and Orit Shaer[4]

1 **Carleton University – Ottawa, CA,** `audrey.girouard@carleton.ca`
2 **University of New Hampshire – Durham, US,** `andrew.kun@unh.edu`
3 **University of Bristol, GB,** `roudauta@gmail.com`
4 **Wellesley College, US,** `oshaer@wellesley.edu`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 19232 "Ubiquitous Computing Education: Why, What, and How". The workshop gathered 26 faculty members and one undergraduate student to discuss the current state of ubiquitous computing education, and how the training and education in this domain should evolve. We provide the motivation for the seminar and an overview of the activities. The outputs of the seminar include laying out the challenges of teaching ubicomp (WHY), proposing a ubicomp curriculum based on various types of students (WHAT) and innovating active learning methods for ubicomp (HOW).

Seminar June 2–7, 2019 – http://www.dagstuhl.de/19232

2012 ACM Subject Classification Human-centered computing → Ubiquitous and mobile computing, Social and professional topics → Computing education

Keywords and phrases active learning, computing education research, education, human computer interaction, ubiquitous computing

Digital Object Identifier 10.3830/DagRep.9.6.26

1 Executive Summary

Audrey Girouard (Carleton University – Ottawa, CA)
Andrew L. Kun (University of New Hampshire – Durham, US)
Anne Roudaut (University of Bristol, GB)
Orit Shaer (Wellesley College, US)

This document summarizes the insights gathered during the seminar. We first provide an overview of the motivation for this seminar before presenting an overview of the activities that occurred during these five days. We then provide a series of outputs that we gathered in addition to the list of abstract provided on the website.

1. Motivation

Interactive systems are becoming increasingly complex and diversified, often comprised of multiple interconnected devices, with many different functionalities. They are slowly merging within our everyday objects. Such systems are becoming ubiquitous. Ubiquitous computing, or ubicomp, is a multidisciplinary field of study that explores the design and implementation of such embedded, networked computing systems. Due to the novel aspect of the technologies involved and the multidisciplinary nature of skills needed to design such systems, teaching and training new innovators in this field are not well addressed through traditional programs and instruction. Consequently, it is important to ask several questions about the training and education needed to help students become valuable members and leaders of ubicomp teams. Three central questions about ubiquitous computing education emerge: why, what and how, with the goal of enhancing ubicomp education through interdisciplinary perspectives:

- WHY is training in ubicomp needed? Is it enough to train experts in narrow domains (e.g. those who can create low-power embedded circuits, or those who can make usable applications), and then bring them together in teams that will tackle ubicomp problems? Or do we need specialized training that targets ubicomp in addition to domain expertise? There is broad consensus that we do need specialized training, but often this argument is based on intuition and anecdotal evidence. We approach this question by first asking: what are the grand challenges that we expect our students to tackle in the world (e.g. privacy, sustainability) by inventing and developing ubicomp solutions? Next, we ask: who can better address the challenges: teams of domain experts, or teams where at least some team members have specialized ubicomp education? Answers to these questions will clearly identify problems that might exist with current ubicomp educational approaches.
- WHAT should constitute training in ubicomp? Once we identify the grand challenges, we need to ask further questions. What are the values, knowledge, and skills we should train students in ubicomp? What are the topics that should be covered? How do these depend on the background of students or their degree program? Answers to these types of questions will allow us to set goals for ubicomp education.
- HOW should we teach and engage a diverse body of students? Once we identify specific goals for ubicomp education, we need to ask ourselves how those goals can be achieved. How does the unique nature of ubicomp challenge the current pedagogical approaches? How can we create new pedagogical approaches for teaching and training in ubiquitous computing? Answers to these types of questions will help create the appropriate tools to reach our ubicomp education goals.

2. Overview of the activities

Our goal was to create a community to support new forms of teaching, training, and learning in ubiquitous computing. Our activities were centered on our main questions:

- Day 1, we explored the WHO and WHY. Each participant presented briefly their research and current teaching, and highlighted what they see are the main challenges for teaching ubicomp in the morning. We then brainstormed and discussed why is is important to rethink the way we teach ubicomp material and what are the grand challenges associated to this change.
- Day 2, we explored the WHAT. In groups, we defined the curriculum for Ubicomp education for different types of students, different degree levels, as well as identified what are the learning goals. One discussion that came up relating to the limits of Ubicomp material, specifically how complex it currently is to define what is ubicomp.
- Day 3, we explored the HOW, and particularly brainstormed about the challenges related to ubiquitous education. Participants generated a list of their current active learning methods or tools and exchanged them in a speed dating fashion with each other.

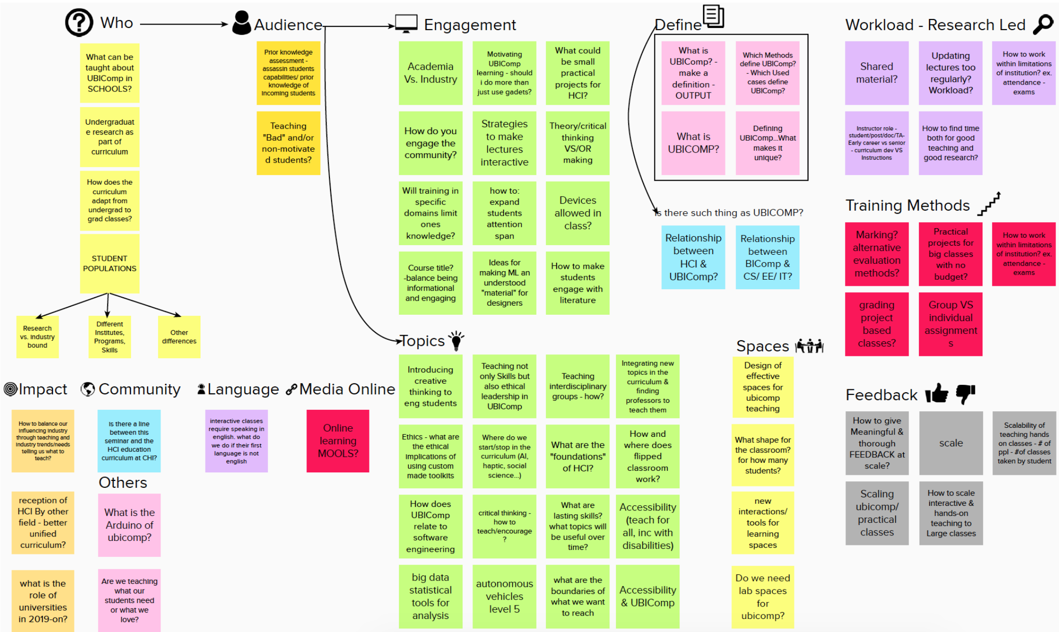

Figure 1 Grand challenges of teaching Ubicomp (from a motivation point of view).

- Day 4, we explored further the HOW. In groups, we developed and experienced new active learning pedagogies on ubiquitous computing topics. We also discussed pedagogies for academic ubicomp programs.
- Day 5, we wrapped up the seminar and plan for concrete actions for the future, in particular, ideas for the next Dagstuhl seminar.

3. The challenges of teaching Ubicomp (WHY)

Figure 1 illustrates the grand challenges of teaching Ubicomp from a motivation point of view. We have identified several themes including (1) who is the audience in terms of diversity, motivation, population; and how (2) these aspects particularly impact their engagement and what methods can we use to better engage with students. We also talked about the difficulty that Ubicomp brings in terms of being a multi-disciplinary field and we highlight the fact that it is difficult to choose (3) which topics should be covered and which ones should not be covered in a particular case. What are the boundaries of Ubicomp? In fact, our discussions highlighted that there is not a clear (4) definition of Ubicomp. We talked about (5) issues with the high workload of both teaching and learning about ubicomp, and how research-led teaching could alleviate some of these issues. We discussed (6) scale issues, i.e. how to teach to a large number of students (and provide feedback) when it seems that certain aspects of Ubicomp teaching (e.g. workshop activities) can only be taught to smaller groups. We pointed out the issues of (7) space and that Ubicomp teaching is based on traditional classroom but also new types of spaces such as workshops, hackerspaces, and maker spaces. Furthermore, we discussed other media types such as (8) online lectures. We also discussed more general topics such as (9) the impact of ubicomp (e.g. on business and industry) and the future of universities and how this relates to ubicomp education.

4. The Ubicomp curriculum (WHAT)

Table 1 illustrates the topics central to Ubicomp Education brainstormed during the seminar. We split the participants (including the organizers) in four groups designing curriculum (standalone lecture or program) for different students (undergraduate UG or postgraduate PG) and technical (Computer Science) or non-technical (Interdisciplinary) background). We wish for this document (that we also plan to put onto our online web platform) to be used as guidelines for teachers in order to provide a better and unified Ubicomp curriculum across different institutions and countries.

5. Existing active learning methods for Ubicomp (HOW)

Figure 2 illustrates the grand challenges of teaching Ubicomp from a method's point of view. This was the result of a brainstorming with participants following the curriculum creation. We found that (1) managing the workload was a theme recurring again (as we also mentioned it in the initial brainstorming in Figure 1). We noted that one difficulty of teaching Ubicomp was (2) the lack of differentiation with other CS, HCI or Design teaching material. We also though this could create issue in (3) attacking certain types of students and that possibly, depending on the demographic, different terminology (Ubicomp, Interactive Systems, Interaction Design etc.) might be used. We raised issues in (4) evaluation and assessment potentially raised by (5) the interdisciplinarity of the community which makes it hard to assess student but also to teach so diverse material. We discussed issue in (6) engaging with students and enforcing skill acquisition (surface vs. deep learning). Finally we also add other issues such as (7) scaling of students, (8) project styles, (9) reaching to real end-users, (10) having input from industry and the (11) format of the lecture (e.g. online). We finally discussed about the issues raised by (12) admin and physical resources.

To build on participants' past and current experience regarding education, we also asked them to share both memorable experiences as well as active learning methods. For the former, we wanted to gather memorable educational moments, anecdotes that stayed with participants long after, as means to both remember the impact that we have on others, as well as get inspiration when designing new activities or methods.

We finally asked participants to share three teaching active learning exercises or methods, ones they currently use in their teaching materials, or ones they experienced in the past. They shared their methods, in a one on one, speed-networking format. In two minutes, they explained one of their ideas to another participant. After the speed-dating, participants placed their basic descriptions on a board and voted for the ones that seemed relevant to their courses. This activity sparked interest for material sharing among participants and ideas on how everyone could implement some of the approaches in their own contexts.

6. Innovative active learning methods for Ubicomp (HOW)

The next main activity focused on generating new educational material that may be difficult to generate, or missing, from a current curriculum. Participants formed six groups, they selected a topic, and investigated new active learning methods as well as initial teaching material related to the topic. The specific topics for each group were selected from topics and challenges highlighted earlier in the seminar. Next, groups formed pairs of groups, and

Table 1 Ubicomp Curriculum by themes, topics and types of students.

Theme	Topic	Course UG Interdisciplinary	Course UG Technical	Course PG Interdisciplinary	Course PG Technical	Course PG *Architecture*	Program PG Technical	Program UG Interdisciplinary
Introduction	Visions of ubicomp	x	x	x	x	x	x	x
	Ubicomp Fundamentals	x	x	x	x	x	x	
	History of computing	x	x		x	condensed		x
	Interfaces	x	x	x		x		
Tech	Hardware skills	basic	x	x (prior knowledge, class, or bootcamp)	x	basic	x (prior knowledge, class, or bootcamp)	x
	Software skills	basic	pre requisite	x (prior knowledge, class, or bootcamp)	x	basic	x (prior knowledge, class, or bootcamp)	x
	Fabrication techniques	basic	x		x	basic since prior knowledge	x	
	Interaction techniques, modalities	basic	x		x	overview	x	
	Electronics	x	x	x	x			
	Sensors	x	x	x				
	Actuators	x	x	x				
	Location tracking		x	x				
	IoT		x	x			x	
	Signal processing				x	use toolkits	elective	
	Network	basic	x	x			x	x
	Activity recognition				x			
	Communication protocols				x			
	System building		x		x			
	Infrastructures				x			
	Appliance design		elective		x			
	Displays						elective	
	Content Awareness			x				
	Testing, certification, ISO				x		x	
Methods / Design	User Centered Design (UCD)		x	possibly pre-requisite	x			
	Sketching + design	x	x	x				x
	Prototyping Methods	x	x	x	x			x
	Evaluation		x	x	x	x	x	x
	Inclusive/accessible design			elective	elective		elective	elective
	Statistics				pre-requisite		x	x
Specific domains	AI/ML	basic	optional	optional			elective	
	Human Augmentation	basic	optional	optional				
	Data Science/Analytics			optional	x		elective	
	Robotics						elective	
	Sustainability					x		x
	Autonomous systems		advanced					
	Entrepreneurship			elective				elective
Implications	Ethical considerations	x	x	x	pre-requisite	x	x	x
	Security	x	x	x	pre-requisite	x	x	x
	Privacy	x	x	x	pre-requisite	x	x	x
Communication	Demonstration of product		x	x	x	x	x	x
	Writing/describing product		x	x	x	x	x	x

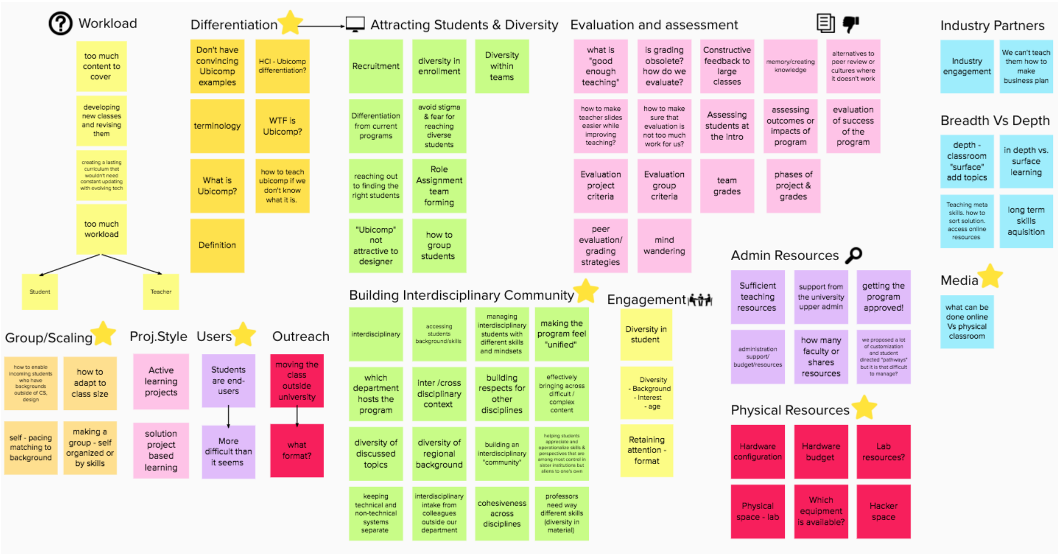

Figure 2 Grand challenges of teaching Ubicomp (from a method point of view). A star corresponds to a challenge specific to Ubicomp while other are general to teaching.

each group tested their content and methods on other group and received feedback, before iterating on their design. Finally, teams presented a summary of their new materials to the group (Table 2).

From the discussions following up the presentation we also noted some actions to do:

- Create a repository or playlist (youtube) or videos that can be used within the community and define what is Ubicomp.
- Ask participants to upload a 2 minutes video of their definition of Ubicomp that can be used in class to show the variety of what people think is Ubicomp.

7. Future Steps

Although this seminar addressed many questions the organisers had originally highlighted, it also opened new exciting directions to explore and new challenges. To start addressing them we identified the main following avenues for future work and future events:

- Follow-up Dagstuhl seminar on writing a textbook
- Follow-up Dagstuhl seminar focussing on the industrial side, e.g. what skills do students need for the society we will built in 5/10/20/50 years?
- Using the website to keep the community alive as well as the access to material, and also create a video channel to create a repository of ubicom examples.
- We also have discussed about 3-4 follow up papers to be written among participants and organizers.

Table 2 Innovative content and activities for Ubicomp education.

Topic	Content	Activities
What is Ubicomp (Ellen, Jakub, Michael)	▪ Multi-media video presenting Ubicomp and that the whole community can use + show examples of real life Ubicomp (or fictional like on movie) ▪ Have a good and bad example of Ubicomp repository video ▪ They showed a video of the construction worker in traditional set up and new (with new digital tool) ▪ Introduce key concepts (e.g. calm technology, Weiser, disappearing computer etc.) ▪ ILO: identify concepts in showed example / explain the example using appropriate terminology ▪ What is Ubicomp " let's people define HCI. It is too big to propose a definition because the edges are so blurry. But this is may be not a bad thing (although students may ask us what it is). ▪ What is ubicomp? May be it is about designing interconnected devices while removing the stress of having so many devices to deal with (from a user point of view)	▪ Make them watch videos and read paper and write a definition of Ubicom (plus sketch) ▪ Definition of terminology (e.g. Cliff notes or cheatsheet) ▪ Give examples of categories of things that can disappear (e.g. task, physical computer) and ask students how would you make it disappear. ▪ Repository of video or youtube playlist?
Electronics (Tim, Brygg, Thomas)	▪ Use three examples ("schools") of building new Ubicomp devices using simple electronics and use these examples all along the course. ▪ Three layers of increasing complexity to focus on are sensing, actuating, computing. ▪ Having a 30 minutes version of this as well.	
Electronic (Nicolai, Jessica, Jeremy, Andrea, Simon	▪ 110 minute lecture about sensors part of a more general course on physical computing ▪ Here focus is on hardware so code is not the focus and is prepared (download and pre-assemble circuit). ▪ Basic of resistors (sensing as a variation of resistance)	▪ Before class: they get material and they assemble and glue together a sensors (plus papers material). They will end up with different sensors ▪ Motivating by showing famous consumer devices and show how it is inside and this is "simple" ▪ Show what's inside a resistor ▪ Find out about the mapping between the sensing value we are getting and the actual response (e.g. acrylic overlays to change the light coming to a sensor) ▪ Collect all broken device and identify things inside.
Interfaces (Anke, Michal, Sylvain, Donald)	▪ Design space to describe the dimension of interacting with Ubicomp. Show example of Ubicomp visions and show where they are in the design space. ▪ It may be good to show old video trying to predict the future to show what things are changing, may be why defining Ubicomp is hard. Are they future of the past?	▪ Activity 1 "understanding possible ubicomp interfaces: gave existing material and let students place them in the design space" (the mock up activity show this may be a good activity). ▪ Activity 2 "Desining with ghosts" (Donald use it already). This is kind of a wizard of oz design where two students are ghosts and actuated things around and the user is blindfolded. The ghost help them through the world by actuating things around. ▪ Bank of videos example
Research methods (Caitlin, Miriam, Vicky)	▪ For interdisciplinary class (45 min) ▪ ILO: identifying research question, making a hypothesis, identifying variables, operational definition and study design.	▪ Give hypothesis example, .e.g "the time of the day will improve your grade" . Ask "Is your hypothesis falsifiable" ? ▪ Activity around what is measured in terms of the dependant variable and how to make sure the measurement fits with the hypothesis. ▪ Watch a short video, come up with a question "does self driving cars cause more accidents" . ▪ Giving example of wrong design or wrong set up and make them find why and where are the problems.
Implication (Aurelien, Eva, Ruzanna)	▪ Use integrated approach to talk about the implication during the entire course rather than in a single lecture ▪ Topics as dimensions to discuss within a diagram they proposed (see Radar diagram in the reading list) ▪ Show the "what our smart devices know about you" (TED talk)	▪ Use case studies to discuss impact of technology or architecture choices on impact topics (from the diagram) as well as other factors. ▪ Exercise with critical review of certain system can be done. ▪ Repository of use case examples (e.g. Amazon dash video, google glass, autonomous cars, Alexa etc.) ▪ "Common thread" with a visual so students know when the implication is discussed during an entire program.

8. Reading list

We collected a reading list that addresses the why, what, and how of ubicomp education, designed for educators.

- Pervasive Computing Education, Audrey Girouard, Andrew L. Kun, Anne Roudaut, Orit Shaer, and Andrew L. Kun, IEEE Pervasive Computing, Oct, 2018.
- Teaching Pervasive Computing in Liberal Arts Colleges, Orit Shaer, and Evan M. Peck IEEE Pervasive Computing, Jul, 2018.
- The Fuzzy and the Techie: Why the Liberal Arts Will Rule the Digital World, Scott Hartley
- Fixing Tech's Ethics Problem Starts in the Classroom, Stephanie Wykstra, The Nation.
- The Pervasive, Embedded, and Mobile Computing Curriculum – Preparing Computer Science Students for the Technology of the Future, Jakob Bardram, 2012
- Bridging the Gap Between Teaching and Research: A Case Study for Engineering & Applied Science, Anne Roudaut, Higher Education Pedagogies 2019
- Krumm, J. (Ed.) Ubiquitous computing fundamentals. CRC Press, 2010 (fairly outdated by now)
- Rowland et al. Designing Connected Products. O'Reilly 2015 (on Design of IoT products, with a broad range of topics ranging from networking aspects, architecture to product design)
- Landay, J. A., & Borriello, G. (2003). Design patterns for ubiquitous computing. *Computer*, *36*(8), 93-95.
- Electronics books from "Make" .
- The radar diagram https://scottwhyoung.com/teaching/information-ethics-privacy-spring-2018/

2 Table of Contents

3 Overview of Talks

3.1 Reflections from Michael Beigl on ubicomp education

Michael Beigl (KIT – Karlsruher Institut für Technologie, DE)

The reflection on what Ubicomp is has taken much of discussion about teaching Ubicomp:

- how much should it include knowledge from other CS and non-CS areas. Is Ubicomp just all of computing (I don't think so)
- in general, does Ubicomp really exist as a topic separate from the rest of CS? (I think so)
- what is the focus, e.g. focusing on systems that are of low (perceived) complexity (calm computing, etc.)? (I think: yes: the core of Ubicomp is about methods (design) and technology to make computers calm and stay in the background)
- Is the Smartphone a legitimate Ubicomp device? (I think yes, but..) What is the difference to Mobile Computing, IoT, etc.? ? (concept vs. technology centered)
- is Ubicomp a subset of HCI? (I don't think so, as e.g. all of tech is missing)
- should there be a development of Ubicomp towards modern times (data-centricity, more complexity)? (yes, but how?)

I see this as positive, as a debate helps shaping an area. And it is always good to reflect on the basics.

3.2 Design patterns for Ubicomp?

Andrea Bianchi (KAIST – Daejeon, KR)

This seminar had the ambitious objective to try to understand the role of Ubiquitous computing in contemporary society and current educational curricula around the world. We started by trying to define what is Ubicomp, and how it differs from HCI education or other multidisciplinary programs. Furthermore, we attempted to define a list of topics that span under the umbrella of Ubicomp. One of the common elements that were highlighted is the fact that Ubicomp does not require the direct attention of users, but it works in the periphery and background ("calm technology"). We further elaborated on this point, and we all seem to agree that it is not a matter of technology, but a matter of user's consciousness and overall context. For example, mobile devices such as smartphones, despite requiring the users' full attention and engagement when used, can also be seen as Ubicomp devices because the usage is transient. For example, using a mobile device to check the time or as an alarm clock are examples of short burst of transient activities that should not be read as isolated independent actions, but in a broader context. Under this light, Ubiquitous computing is not technology driven, but rather depends on the users' consciousness about the activity at hand (i.e. a "background" task when contextualized in light of the "broader" task at hand).

The second half of the seminar is mainly focused on trying to determine the means (HOW) and ultimate motivations (WHY) for teaching Ubicomp in universities. Participants highlighted that the multidisciplinary and generalist competencies usually required in current Ubicomp classes might not be enough for real world products (where instead specialized multidisciplinary teams work together). Furthermore, we can also assume that many of these competences will become more common places for future generations, or might even disappear. The question hence becomes again WHAT type of knowledge Ubicomp can additionally provide to a CS or Design curriculum, or, in other words, what general knowledge will remain in the students considering that specific technologies might become obsolete or common knowledge over time. One of the suggestions that seems to be shared among many participants was that perhaps Ubicomp educators should attempt to distill domain specific knowledge using design-patterns. As many professional fields such as software engineering, interface design and security, attempt to transfer knowledge by building unit-blocks of knowledge in the shape of patterns, so perhaps Ubicomp education will require the same level of abstraction – patterns that go beyond technology, but are specific enough that can be implemented using the current available tools and technology. I personally think that defying these patterns will be next big challenge of the field.

3.3 Teaching Ubicomp in different countries and for different students

Anke Brock (ENAC – Toulouse, FR)

We were lucky to have participants from many different countries and universities across the world in our Dagstuhl seminar 19232. One interesting observation was then how teaching environments and situations vary across these different countries and institutions. This may impact the number of hours which could be attributed to a new Ubicomp course or program, the availability of hackerspaces and hardware, or the background of students (e.g. their prior training). A new Ubicomp course or program would need to be adapted to all these factors. In our working group, we first worked on designing a new Ubicomp course for computer science students, which could be integrated into an existing program (for instance in my case the Master for HCI in Toulouse, France, http://www.masterihm.fr/). Since we aimed at technical students, we expected students to meet certain prerequisites (e.g. programming skills, software development methods, statistics). Our course would then encompass lectures and exercises on topics such as concepts of ubicomp; methods; sensing, data & communication; hardware; interaction techniques; and system. We identified skills and knowledge that should be taught in parallel in general HCI or CS lectures (e.g. user-centred design). We agreed that current students would enter a Ubicomp course without necessarily having knowledge on electronics (e.g. Arduino), but that it might become prerequisite knowledge for students in 10 years from now (similar to the increase of CS teaching in high schools which we currently observe). An interesting exercise was to adapt this course to a different student population. In our case we discussed how such a lecture could be adapted for architecture students who would be interested in learning ubicomp for the design of smartphones. Such type of students would look at ubicomp from a different perspective than CS or HCI students: rather than learning how to build systems from scratch, they would be interested in learning how to make use of existing systems and toolkits. As a consequence, the focus of the lectures would shift from a technical perspective to a more applied perspective. On the other hand, topics such as sustainability would gain more importance, since architects are already trained in considering aspects such as energy consumption in the design of spaces. To sum up, the design of a Ubicomp class or program needs to be adapted to the specificities of each university. However, this Dagstuhl seminar hopefully lays the foundations which will help us implement such a course or program at our own institutions.

3.4 Reflections from Jessica Cauchard on ubicomp education

Jessica Cauchard (Ben Gurion University – Beer Sheva, IL)

While the discussions started around what and how to teach Ubiquitous Computing, we realized that we couldn't fully define how Ubiquitous Computing differs from Human-Computer Interaction education. In addition, we believe that within the next decade or so, Ubiquitous Computing will become the norm in terms of computing. As such, what are the skills required and how large is the scope of university education in the field? One of the discussion topics was around whether data science should be part of Ubicomp teaching curriculum or not. We also wondered how the background of students will change within the next 5,10 and 20 years and how early their education in the field will start.

Following these discussions, we started establishing curriculums for both undergraduate and graduate programs in Ubiquitous Computing. The programs we focused on were targeted for both technical or multidisciplinary students. We found that the programs that were designed by the different groups ended up with many similarities, all including a large range of topics from software and hardware engineering to design and creative skills. The undergraduate programs had a focus on practical skills and industry placements while graduate programs had an emphasis on research skills and research methods. We also had discussions around institutional support and constraints for courses, such as budget for class equipment. When designing classes, most classes comprised active learning activities for students to engage with the content and better understand the concepts.

3.5 Interactive Socio-Technical Systems (or what's next for Ubicomp)

Ruzanna Chitchyan (University of Bristol, GB)

URL https://energysystems.blogs.bristol.ac.uk/author/r-chitchyan/

Can a single curriculum can deliver teaching and training for all knowledge and skills necessary to deliver the newly emerging generation of Ubicomp systems? Let's consider that that to get driverless cars (as an example system) to "disappear into the fabric of everyday" (that is to become ubiquitous) we need to draw on skills and knowledge in: mechanical, electrical, electronics, and software engineering, networking and telecommunications, as well as, ergonomics, aerodynamics, product design, HCI, psychology, cultural context, regulations and laws, to name a few (not even mentioning aerospace related issues for getting the GPS systems in place).

It is then difficult to see how a single curriculum for a Ubicomp systems can be arrived at. Neither do we try to get only Ubicomp graduates to work on driverless car systems today. In practice, when developing and deploying such ubiquitous solutions, teams of specialists collaborate to draw in the required skills and knowledge.

Then maybe we could agree that **Future-Focused Ubicomp Curriculum** does not limit itself to Hardware and Software Prototyping, Programming and Networking, and Creativity modules, but is taught as a **programme where specialists** in Software Engineering, Networking, Electrical Engineering, HCI, Human Anthropology, Psychology, Business, Law, and other disciplines textbfwork in teams to build large, complex, situated, usable systems that become the infrastructure of tomorrow. In other words, the learning to build Ubicomp (or to be more precise, Interactive Socio-Technical Systems) is not accomplished through graduating from a specific degree, but by "graduating" through working on collaborative *interdisciplinary projects with focus on interaction of human (both individual and societal) and technical aspects.* In other words, I suggest that such projects should form the "heart" of the curriculum on Ubicomp/ Interactive Socio-Technical Systems, while the specific curricula would continue to deliver HCI, Communications, Software Engineering, Law, Phycology and other similar modules, depending on the flavour of the courses provided.

3.6 Situating UbiComp Education within or distinct from HCI?

Jeremy Cooperstock (McGill University – Montreal, CA)

Ubiquitous computing education faces a hurdle of differentiating itself from that of general HCI education, and perhaps more importantly, demonstrating its relevance and importance to society at large. Despite the clear definition of Ubicomp as provided almost thirty years ago by Weiser, there is no clear consensus as to what topics fit and which do not, within a course or program that is not primarily overlapping with HCI. As a result, despite a high degree of agreement on topics of relevance, as put forward by the various workshop participants, it remains uncertain as to how this subject will compete with established HCI programs. Whether this implies that usability concerns for handheld computing (and its future incarnations) will remain dominant over the design of calm and "invisible" technology, in particular, in the teaching of future generations of students, remains to be seen.

3.7 Creating Magic– reflections on ubiquitous computing

Ellen Yi-Luen Do (University of Colorado – Boulder, US)

Do you ever wonder what would the future be like? What kind of technologies would people use in their everyday lives? What kind of the reality of ubiquitous computing would be achieved in the 21st century? Instead of singing the cute song "Que Sera, Sera. Whatever will be, will be. The future's not ours to see", we are fortunate to have Alan Kay's suggestion that "The best way to predict the future is to invent it." It's so nice to be able to invent the future we will live in! Then, what kind of future shall we be inventing? How do we prepare our young generations to engage in meaningful and useful ubiquitous computing?

Weiser's vision for ubiquitous computing is about calm technology, in which the technology disappears into the world we live in and become peripheral, and put us at home. This week's Dagstuhl Seminar 19232 – Ubiquitous Computing Education: Why, What, and How created a unique opportunity for a community of scholars and professionals gather together to brainstorm and exchange ideas about ubicomp education and the type of curriculum that would be needed for inventors and designers of future interactive systems.

With various breakout sessions in thematic topics, we asked questions about what is Ubiquitous Computing, and whether it's useful or even necessary to still use the term, or should the term of Human-Computer Interaction (HCI) be the defining domain? How, then, do we differentiate these with the fields of information and the study, creation and analysis of media? Shall we instead use the term interactive systems? There is no one easy answer. We will just have to evolve the field. Of course each name comes with its own promises and baggage, as well as the persuasive powers to attract different kind of audiences. I am delighted with the active discussions and the questioning about ubicomp education, and pleased with the enthusiasm. Meanwhile, let's not forget the suggestion from the futurist, inventor, and science fiction writer Arthur Clark that "Any sufficiently advanced technology is indistinguishable from magic." In our design of the curriculum we certainly would like to empower students with technical competency, but also the creative spark and the power of play in creating magic!

3.8 Reflections on the need to discuss ubiquitous computing education

Audrey Girouard (Carleton University – Ottawa, CA)

The ubiquitous computing seminar offered attendees an opportunity to apply their research expertise to improve the delivery of the topic to undergraduate and graduate students. This was a fairly unique opportunity: education is usually discussed internally within institutions, but rarely globally, other than in specialized conferences. Yet, education is at the core centre of our mission as researchers and faculty members. Teaching students the right topic using appropriate and engaging active learning methods can be challenging. This seminar stem from the organizers' interest to learn from their colleagues' various perspectives and experiences, to create a large community of teachers that can learn from one another. The seminar grew to become a place to discuss the vision of teaching ubiquitous computing, establishing the general topics to teach a well-rounded class, as well as proposing how to put that in action.

The seminar reflected on vision of teaching in the 5 to 20-year horizon, as well as design new curriculum based on how the field and teaching methodologies are evolving. A core challenge that emerged rapidly and recurring throughout is the difficulty of defining what is ubiquitous computing, what exactly it englobes, what are the salient examples that define it. Very few of the 28 participants teach a class explicitly named ubiquitous computing, though many have related topics, or include it as a topic instead of a focus. Another interesting observation was the necessity to distinguish general teaching and learning-related discussions to those that are ubiquitous computing specific. Ubicomp focused challenges included the necessary to differentiation between ubicomp and other related topics (e.g. general HCI), physical resources (lab space, hardware budget, etc.), and interdisciplinarity in the topic. Engaging though general topics included teaching workload, student recruitment and student engagement in classrooms.

Overall, this was a fruitful and engaging seminar with discussion and materials. It will lead me to update my teaching materials shortly and improve the delivery of ubiquitous computing to my students.

3.9 Reflections from Eva Hornecker on ubicomp education

Eva Hornecker (Bauhaus-Universität Weimar, DE)

There was a lot of comments on what's the difference between HCI and Ubicomp. For me, the UbiComp we are discussing here (which is the more HCI'ish section of UbiComp) is a subset of HCI, a specialization, with some overlap to the more technical areas. It is specialized in terms of the types of technologies involved, as these have some form of sensing technology (often also acting back on the world), and interact with or relate with the outside world and objects (unlike general apps, which might just live on a phone – and even if used in a mobile situation, might not interact with the outside world). So this requires an understanding of the technology and its embedding in the world, and what follows from this. Many of the challenges with teaching we discuss appear to be of a more general nature for HCI and applied topics, still it is good to have an opportunity to discuss these. Devising a curriculum for UbiComp was interesting, in particular thinking about what a full degree could or should involve. The collection of teaching materials that was prepared on Thursday will be incredibly useful, and even though I'm already teaching Ubicomp, there are a lot of new ideas, in particular for interactive elements and activities for some of the Ubicomp core topics.

One of the reasons I like teaching UbiComp is that it lends itself to a holistic, systemic approach, which discusses applications as well as societal and ethical concerns (that partially may be specific to UbiComp, but should be considered in a technical degree education anyway). In UbiComp teaching, we need to discuss technology, hardware/sensors, system architectures, discuss applications, etc., and here discussions of impact (ethics, ecology, privacy) can be integrated easily ad hoc. This holistic approach also makes the class accessible and interesting both for more technical students and for students with a more human science or design interest.

3.10 Integrating Ubicomp in undergraduate education

Miriam Konkel (Clemson University, US)

In the context of a workshop discussing Ubicomp in Education, we discovered that participants have different conceptions of the nature of Ubicomp. From the early 1990s context, there are well-known descriptions by Weiser. However, as some pragmatic understandings of ubiquitous computing are now deeply woven throughout our daily lives, it raises the question of broader present-day definitions and academic implications. We also realized the prospective value in characterizing the boundaries between Human Computer Interaction and Ubicomp. For example, is Ubicomp a part of Human Computer Interaction, or are they two separate disciplines with considerable overlap? We are aware that our student population is diverse. We see the need for transdisciplinary programs as well as technology-centered programs. Our group focused on program development for undergraduate students who are pursuing an interdisciplinary Ubicomp minor. Our focus group itself was interdisciplinary with domain experts ranging from computer science/engineering, design, and psychology. For an interdisciplinary program, we anticipate a similar spectrum of students. Students might receive training an Ubicomp overview course that covers the past, present, and future of Ubicomp. In this course students would also be introduced to ethics and the need to consider inclusiveness for the development and design of prototypes. These topics will be woven throughout other courses. We anticipate that students will be able to choose from a subset of courses ranging from programming, computer engineering, psychology, research methods, statistics, and object design.

3.11 Ubicomp as a vehicle for teaching students to generate new knowledge

Andrew Kun (University of New Hampshire – Durham, US)

Having designed and taught two different ubicomp courses, I'm very interested in looking ahead to the next decade or more, and asking: what is the future of ubicomp education? This is the reason I was excited to co-organize Dagstuhl workshop 19232 "Ubiquitous Computing Education: Why, What, and How."

I believe that one central issue for ubicomp education is to help train students to successfully generate new knowledge. Ubicomp is a field that is advancing quickly, and many of the specific tools of generating new knowledge become obsolete quickly [1]. For example, the programming tools, and the tools for creating hardware change rapidly. Students need to keep up with this change in order to be successful in their future workplaces. However, the need to keep students up-to-date on the specific tools of the trade should not prevent us, educators, from also training them in the general tools of generating new knowledge. The general tools are described by Karl Popper's model of conjecture and refutation, in which scientific theories are built through proposing hypotheses which are subsequently empirically tested and either supported or refuted. Steven Pinker describes these general tools in terms of Bayesian reasoning: certain hypotheses are given a baseline probability based on prior knowledge, and this probability is adjusted based on empirical tests. Importantly, these skills are not only useful in scientific research in academia. As the recent work of Eric Ries demonstrates, these general tools are also critically important for productive research and development in industry.

Fortunately. Ubicomp education presents educators with an excellent opportunity to train students in both the specific, and the general tools of generating new knowledge. This is because very often ubicomp education focuses on creating artifacts. This process requires specific tools – students need to know how to program the latest gadgets, and how to use the latest hardware prototyping tools. However, students also need to create experiments in which they test their creations and assess their viability in different situations. Educators should seize these opportunities and help students gain an appreciation for, as well as training in, generating research problems, hypotheses, and approaches for testing hypotheses.

References

1 Girouard, A., Kun, A. L., Roudaut, A., and Shaer, O. Girouard, A., Kun, A. L., Roudaut, A., & Shaer, O. *Pervasive Computing Education*. IEEE Pervasive Computing, 17(4), 9-12, (2018).

2 Kun, A. L. Kun, A. L. *Reader and Teacher: Fourteen Books that can Inspire Teaching in Pervasive Computing (and Beyond)*. IEEE Pervasive Computing, 18(2), (2019).

3.12 Is Ubicomp here?

Sylvain Malacria (INRIA Lille, FR)

Mark Weiser, the father of ubiquitous computing, predicted in 1999 a future where everything will be able to sense human input and augment its appearance with digital information: essentially computation will be fully embedded into all aspects of the real world and the transition between the physical and the virtual would become seamless for the user. Today is becoming this future, with the spread of interconnected devices not only in our pockets (smartphones, smartwatches, etc.) but also in our environment (smart buildings, connected bus stops). Interestingly enough, the fact that ubiquitous computing has become so "real" has also made the notion of ubiquitous computer itself less "clear", in terms of which notions it encompasses and how they should be transferred.

During this seminar, several participants agreed that ubiquitous computing is a rich interdisciplinary domain, hard to define, and whom boundaries are hard to identify. It has been clear, however, that designing ubicomp systems requires a variety of skills usually instructed in various disciplines. After discussing what the typical student in ubicomp will be in 5,10 and 20 years, this seminar tried to define typical programs and courses for teaching ubicomp for both technical or interdisciplinary groups of students.

3.13 From Weiser to Now: The Search for Ubicomp

Nicolai Marquardt (University College London, GB)

Many inspiring conversations during the seminar brought up critical aspects to consider when putting together a possible ubicomp curriculum. First, it is difficult to clearly outline the scope and boundaries of what ubicomp includes. Individual definitions of what ubicomp is was quite diverse across all participants. Overlap and distinction to other technical HCI courses/modules is important to best position the strengths of a ubicomp focused course. Second, we might need to better understand the future career paths of the students, as their jobs in industry will likely not be framed around ubicomp and the particular ubicomp use cases, but cover broader areas across CS, automation and autonomous and networked systems (which is different to many other CS specialisations, such as AI/Machine Learning or Financial Computing, where career trajectories are more obvious). Third, it was discussed that we need a better differentiation between the fundamental principles vs. the applied ubicomp teaching. For example, what techniques and approaches can we teach our students that are still relevant in 10 or 20 years. Technologies, programming languages and tools we use today are outdated very quickly, and so the question is what are the skills and techniques that last. Related to this, it was mentioned to possibly consider creating collections of design patterns focusing on the ubicomp context. Last, we probably need to revisit Weiser's vision of ubiquitous computing (with all its qualities of calm computing, peripheral interaction, foreground/background interaction, situated in people's environments, and so on) and ask how much of that vision and definition of what ubicomp is still applies, should be changed, refined or broadened.

3.14 What is Ubicomp?: A Student's Perspective

Amanda McLeod (Carleton University – Ottawa, CA)

We started off the 19232 seminar with the "Why?" of ubiquitous computation. I immediately began seeing similarities with the current Information Technology (IT) and Human Computer Interaction (HCI) terminology and curriculum that have, and are currently being taught to me in the Interactive Multimedia and Design program at Carleton University. On the second day of the seminar, we explored the "What?", as in what would be taught in a new Ubiquitous computational education curriculum? Teams explored different themes and created course outlines in which we would need to cover multiple topics in order to confidently say a student has been educated in Ubicomp, whoever, all this exercise did for me was strongly establishing the parallel between current educational topics that have previously been seen in my past and present courses, for example, ethics, hardware, software, prototyping, statistics, and many more.

Following multiple discussions with the seminar participants, it appeared as though the established definition of ubicomp continued to blur. Multiple perspectives shaped the name to be about a higher complexity item. As discussions went on, the narrative of ubicomp education rapidly formed a Venn diagram with the HCI world. Both poles of the diagram overlapped greatly in terms of curriculum presentation and thoughts on what the curriculum should become. These repeating curriculum terms posed the questions of "What is ubicomp? What is it not?", "what is HCI? What is it not?". Do both of these terms occupy two completely separate worlds or do they morph into one under the same umbrella?

Additionally, while developing curriculums and methods of delivery, a great portion of the participants shared ways in which they capture their audience's attention. This made me realize that for the past 3 years of my undergrad, I've been the subject of teachers tactics that attempted to get me more involved in class learning and discussions. During this same conversation, the subject of teaching evaluations were brought to the attention of the participants and their major impact on teachers opportunity for promotion as well as salary. This made me reflect on the existing teaching evaluations and how they are presented to the students and how some questions are quite frankly irrelevant to how well the professor taught the class. "Did the teacher speak loudly and clearly?", "Did they end class promptly and on time?". These existing teaching evaluation questions prompted me to question why we don't cater these questions to specific classes in the same way we cater class content to particular classes. For example, if we question a student on if they believe they were sufficiently taught about specific learning outcomes from the course outlines, we would be able to sufficiently evaluate a teacher.

3.15 Reflections from Donald McMillan on ubicomp education

Donald McMillan (Stockholm University, SE)

One discussion on the future of Ubicomp education that resonated was on the future students, their needs, and their abilities. As any technology spreads throughout society, it is appropriated, used, disassembled, and taught by and to a much wider audience than the technical experts, which sparked its creation. As Ubicomp makes its way through transport, architecture, and into K-12 education as part of the infrastructural facilities we can expect it to follow this trajectory (seen, for example, in web technology) of being taught as something to use, then as something to understand, and finally as something to manipulate and create (with or through). This pushes the underlying goal of teaching ubiquitous computing from the practical, technical aspects to more theoretical, abstract, and universally applicable principles – ones that encompass design, ethics, societal impact, as well as technology. This comes hand in hand with the technology we are talking about following the students into this future classroom (in whatever form it will take), and through their lives. This will necessarily change the ability and understanding of the students. In a manner akin to smartphones providing communication and search functionality being taken advantage of in current educational contexts, in the future the trace data from a host of ubicomp systems, possible social and cognitive enhancements, and access to the outputs of these in an interconnected and social manner provide a fascinating glimpse of a possible future of education not only of Ubiquitous Computing, but through and of Ubiquitous Computing.

3.16 Reflections from Tim Merritt on ubicomp education

Timothy Merritt (Aalborg University, DK)

Ubiquitous Computing is a topic that is defined in various ways from researchers, artists, and practitioners from diverse perspectives. It became apparent to me that many agree that Weiser's 1991 paper,"The Computer for the 21st Century", is a good starting point and a way to open the discussion, yet from there, current examples and aims of related research differ dramatically. From a technical perspective, challenges arise when we try to define what is within scope and out of scope for Ubicomp. Is the disappearing computer the main intersection about which most agree?

Weiser discusses location and scale as crucial to the future of ubiquitous computing and paints a picture of a world free from common struggles. Actuated desks mean we no longer misplace papers in the office, deeply connected texts such as physical maps and electronic information, etc. While this vision is optimistic, there are, of course, implementation issues and ethical issues of connecting and actuating the physical world. Further, how should we teach for those who might research or work in related practice? Trends in hardware becoming cheaper, smaller, and easier to program has made "making" more accessible to many more people. Similarly, software is easier to create active environments utilizing toolkits and high-level languages.

Almost 30 years after Weiser's seminal paper, there are some good examples of connected devices and seamless interactions, yet so many experiences with technology are problematic. Voice assistants are often awkward and require abrupt command and terse speech. This makes the technology conspicuous, yet in a very negative way. Aside from these clumsy speech interactions that disrupt social situations, they currently provide very limited opportunities for end-user programming and customization. The calm and smooth world is still elusive. Another technology that we find in everyday life, the smartphone has become pervasive and helpful in so many ways, but when we consider ways in which it fits into our social experience, it too is clumsy and problematic–this multi-purpose device has become almost a wearable companion device, ready at every moment to be helpful in so many ways, wayfinding, controlling home lighting, buying things, banking, socializing, almost everything is controlled through the small screen. Is the mobile phone the cursor for interacting in the world? Will we break free from the small screen and find more embedded displays woven into the physical world? Is the text and information provided on the small screen akin to Weiser's example of the wrapper on a piece of candy? Ubiquitous computing might contribute to many of the concrete challenges faced in world–some of the envisioned smart environments and calm computing experiences have come to fruition and so many remain unsolved. When teaching ubiquitous computing, I hope we can inspire future generations to be creative, innovative, and at the same time, ethically responsible as we involve users to engage with these challenges and design the future together.

3.17 Ubiquitous computing: Accepting a fuzzy field

Caitlin Mills (University of New Hampshire – Durham, GB)

One of the key issues that came up this week was the definition, validity, and purpose of ubicomp as a field. I was left wondering, however, if the definition aspect is necessarily the most important question to tackle at this point. From a philosophical perspective on scientific fields, it is expected that fields will have 'fuzzy' edges, and the very fact that we are organized around a central theme here argues for ubicomp as a valid scientific field (Casavall & Fang, 2015). This perspective further holds that the social aspect of the field is of utmost importance; current (self-organized) members of the field shape what questions should be pursued, what are the appropriate methods, what are the ethical standards, and perhaps most importantly, who will join the field in the future? These questions seem particularly important in the context of education since self-identified ubicomp members will undoubtedly impact the future of the field through the classroom.

Through breakout groups, we have identified the core areas that should be part of ubicomp curriculums, and basic interdisciplinary skills that students would be expected to have. We also discussed potential pathways for how such curriculums/courses may be adaptable based on prior knowledge and expertise. Finally, a substantial time was devoted to addressing topics related to ethics (appropriate methods), diversity, and inclusivity within a ubicomp education– important issues for the future of ubicomp membership.

References

1 Casadevall, A., & Fang, F. C. *Field science—the nature and utility of scientific fields.*(2015).

3.18 Reflections from Simon Perrault on ubicomp education

Simon Perrault (Singapore University of Technology and Design, SG)

During the seminar, we found out that we could not exactly agree on the definition of what ubiquitous computing is. As pointed out by a participant, it was sometimes easier for us to identify something that is not UbiComp or not part of UbiComp. Another emerging thought was that UbiComp and Computing would simply become near-synonyms in the near-future. After discussing modules/programs design, all the participants seemed to reach a consensus on the following notions being part of a successful UbiComp training: hardware and software prototyping, networking, HCI. We also identified related topics that would be relevant to be part of a UbiComp teaching, such as machine learning, data science. UbiComp teaching, by nature, ends up being rather multidisciplinary, and with some changes of focus, could be a suitable topic to teach in many different majors (CS, Engineering, Design, Architecture, Social Sciences and even Business or Law).

3.19 Leveraging the multidisciplinary aspects of Ubicomp

Thomas Pietrzak (INRIA Lille, FR)

One of the many challenges in teaching, studying and conducting research in Ubicomp is the multidisciplinary approach it requires. Similarly to HCI, it requires a broad range of skills that makes it challenging, but also exciting. Most of the participants have a computer science or engineering background. However we also have to teach topics related to design, psychology, mathematics, physics, and so forth. Putting together a curriculum or even a course is a challenge. Firstly because of the long list of topics that have to be covered. But also due to the diversity of students previous knowledge and experience. We designed a curriculum and course that can be adapted to students background (tech or not), and level (undergraduate, graduate). We believe these efforts will contribute to the promotion of Ubicomp, and raise awareness about it among students, and the society in general.

3.20 Reflections from Michal Rinott on ubicomp education

Michal Rinott (SHENKAR – Engineering. Design. Art – Ramat-Gan, IL)

The workshop is a good opportunity to understand the perspectives of different participants on Ubicomp, be it a more system-centered outlook or rather human-centered one; an extreme ("in Ubicomp all interfaces disappear") or a broader perspective on ubicomp interactions. It seems that the lack of a specific and definitive definition is a necessity for maintaining the breadth and wealth of the topic, especially in the context of the workshop. However when planning specific learning experiences there emerges a need for more specificity. Another interesting issue concerns the best conditions for teaching and learning Ubicomp. Specifically, how do we create a physical space that is optimized for learning and research? How does this relate to the ability to scale teaching and learning beyond the people that can work together in a hands-on session? The hybrid learning model of non-synchronous learning of certain materials, coupled with co-located opportunities for group work and interdisciplinary collaboration seems promising. The space in this case should contain tools and methods for supporting this type of meeting. It should answer both specific ubi-comp needs (sketching, prototyping, implementing) and more general tools for collaboration (dynamic structures for group learning, presentation, exhibition). An opportunity arises for the space to implement a ubicomp protocol and, through data and interfaces, provide insights about collaboration within it, and tools to improve learning and functioning within it. The topic of hybrid learning spaces and data is being explored recently by colleagues of mine, for example in this workshop, see open CFP!

3.21 Reflections from Anne Roudaut on ubicomp education

Anne Roudaut (University of Bristol, GB)

The goal of the seminar was to rethink the form of current ubicomp teaching and the role of universities in education in 5/10/20 years. A very insightful idea was to change the current traditional form of courses to a more modular one where students can "pick and choose" modules depending on their existing skills and based on what they want to major in (and consequently shape their future career). This could be done through some kind of visual infographics representing pathways that students took. In parallel pathways that industries need could be assessed every few years to better understand what are the skills needed and help the students choose their pathways. Doing so would require tackling some new challenges such as timetabling but also the need to access skills of students entering the program (and how to avoid students taking the "easier" modules). Such model would also be competing with online courses that are rather following this modular approach (e.g. Khan academy) without necessarily producing accredited diploma.

3.22 Strong Disciplinary Skills are Key – Ubicomp Teaching is then the Multidisciplinary and Practical Application of these Skills in Context of a Realistic Problem

Albrecht Schmidt (LMU München, DE)

The vision of ubiquitous computing has inspired a generation of computer scientists to make computing part of everyday life. At the time Mark Weiser [1] wrote about the computer for the 21st century desktop was dominant and the vision of networked computers everywhere was still far out. As technologies have progressed, we have seen that networked mobile computing has become ubiquitous and is an essential part in everyday life. Sensing and actuation in the infrastructure as well as in devices is common and people constantly used networked services that. However when we think of a parking garage that has sensing to keep track of which parking spaces are occupied and make this online accessible, when the navigation service adapts to sensor information from cars, and when the tools are recording their use automatically in a database we do not talk about ubiquitous computing. Once the technology is woven into the fabric of everyday life, we consider it normal and rely on it. We are currently at a point where hardware cost is becoming a minor factor in comparison to the cost as of software development, system deployment and maintenance. A wireless computer (e.g. ESP8266) is available in 2019 for less than 2€. The limiting factor is our ability to create meaningful systems and applications that have utility and the challenges are all along the life cycle [3]. One further key aspect is how to design the interaction with ubiquitous systems [2] where there is little systematic understanding. Can we improve our ability to creating ubicomp systems through teaching? What will people need to know in order to envision, design, prototype, create, deploy, and study ubiquitous computing systems and applications? At this point I think the field of ubiquitous computing is so wide that it is hard to imagine to set-up a course on ubicomp. I think the best approach is to teach solid foundations in one subject (like computer science) and only add a high-level course on the concept and visions of ubiquitous computing (e.g. a seminar). Additionally having students doing practical projects in ubicomp will help to apply knowledge and deepen it in the field. The serious learning of how to create ubicomp applications will be hands-on in multidisciplinary teams based on strong disciplinary skills.

References

1 Weiser, Mark. Weiser, Mark. *The Computer for the 21 st Century.*. Scientific american 265.3 (1991): 94-105.
2 Schmidt, Albrecht, et al. Schmidt, Albrecht, et al. *Interacting with 21st-century computers.*. IEEE Pervasive Computing 11.1 (2011): 22-31.
3 Broy, Manfred, and Albrecht Schmidt. Broy, Manfred, and Albrecht Schmidt. *Challenges in engineering cyber-physical systems.*.Computer 47.2 (2014): 70-72.

3.23 Reflections from Oliver Schneider on ubicomp education

Oliver Schneider (University of Waterloo, CA)

We continue to struggle with a definition of what Ubicomp education should contain. I have tried to answer three questions to help with that.

What is Ubicomp? → Physical interactive systems, but that doesn't help us define its scope. Physical interactive systems require sensing and a response, either to the user or other stakeholders. It overlaps with Human-Computer Interaction (people and interactive systems). Because these systems are physical, the field involves hardware; because these systems are computational, they involve software; because it involves interacting with people, human-facing fields are involved.

How can we teach it? → Education is difficult because it is interdisciplinary and involves physical systems. Since Ubicomp involves hardware, software, and people, it necessitates different disciplines working together. Educators need to position students' needs within each of these subtopics, and bring them together in a high-level framework. Although education continues to scale through online methods, we still need physical infrastructure for prototyping. Educators will need to decide what students can learn online and what they need to learn in-person. Ubicomp requires physical systems and expertise in multiple disciplines.

What might help us define Ubicomp education? → Don't ask what Ubicomp is, but rather what it is not. Ubicomp is not app design; it is not VR headsets; it is not pure automation; it is not a more precise machine learning classifier. I think that looking at the complementary set will be more fruitful for positioning and defining the field.

3.24 Reflections from Orit Shaer on ubicomp education

Orit Shaer (Wellesley College, US)

The notion of ubicomp has inspired my research and teaching throughout my career – offering the promise of augmenting our everyday environment through the integration of interconnected computational devices. At the core of ubicomp is the idea that technology becomes invisible and accessible anytime and anywhere. Yet, invisible technology, has visible impacts on individuals, communities, and society. While ubicomp presents opportunities for improving our work and wellbeing, it has potential for misuse, in particular, compromising people's privacy and security.

How do we engage students in considering both the promise and challenges of ubicomp? How do we prepare them to become ethical citizens and leaders in an era of ubiquitous computing and rapidly increasing automation? What are the core conceptual and technical skills required for training students to become ethical ubicomp innovators and developers? What methods are effective for teaching complex and interdisciplinary topics to different audiences? The goal of this seminar was to address these questions through systematic investigation of the Why, What, and How of ubicomp education.

The seminar provided valuable and rare opportunity for interaction and exchange of ideas about teaching with colleagues who are experts in the field. While we often meet to discuss our research, it is rare for us to engage in deep and prolonged discussions about teaching, yet we are all educators in addition to researchers. Through small group discussions and hands-on activities, we questioned the definition of ubicomp, examined the purpose and challenges of ubicomp educations, debated the essential content of ubicomp courses for different audiences, exchanged and develop new methods for effective (and engaging) teaching.

Following the seminar, I feel inspired and motivated to reflect on my own teaching, rethink the concepts and ways that I teach, and engage with my students in new ways. The outcomes of the seminar, which include a living curriculum document, reading list, shared educational activities, will be extremely helpful in this process. I also believe that the connections and network of committed ubicomp educators that we formed here has a critical role in shaping how we train the next generation of ubicomp innovators and leaders.

3.25 Scoping Ubicomp for interdisciplinary students

Jakub Sypniewski (Universität Salzburg, AT)

Designing an Ubicomp curriculum brings up a number of connected topics, not only connected with the educational side of things, but also the scope and definition of what Ubiquitous Computing is. One thing that the majority seem to agree on is that Ubicomp requires cross-disciplinary knowledge for both teachers and practitioners. This realisation connects to the already mentioned scope, especially when looking at different student profiles in Ubicomp courses or programs, meaning how technical the curriculum has to be to provide the students with enough understanding or skills related with the computing part of Ubicomp, and how much of the design or concept related to Social Science should be taught not to alienate technically oriented students? The possible way of addressing this challenge would be to gauge students' skills and interests at the beginning of the course or program and adjust the depth to which the topics will be taught. The assessment of students and adjustment of the curriculum might increase workload for the faculty, but is crucial for avoiding teaching unnecessary or out of scope topics.

3.26 Reflections from Aurélien Tabard on ubicomp education

Aurélien Tabard (Université de Lyon, FR)

From developing as an academic research area over more than two decades, Ubicomp vision has made its way into the fabric of our everyday lives. Embedded and inexpensive devices are now spread and networked throughout our environment, from wearables to smart buildings, libraries, hospitals, or urban infrastructures. One could argue that in a way, similar to AI, Ubicomp is defined by a vision rather than a set of principles, and as principles and technologies mature and become broadly adopted, they are not considered as Ubicomp anymore.

Designing and developing Ubicomp systems requires to bring in a variety of skills that have been traditionally mastered by various displicines. The precise blend of skills involved is what makes Ubicomp unique. And the challenge lies as much in the mastery as in the way they are articulated. Broadly speaking, this means introducing sensing and embedded systems specialists to HCI and design, and designers to electronics but also networking and computing systems.

In terms of design approach and practice, I find teaching Ubicomp because 1. It pushes to go beyond the notion of supporting a user accomplishing a task, to consider blending more deeply into existing human activities. 2. It pushes to think of systems at another scale than users or collaborative groups, e.g. designing for bodies, places or cities.

Because of the various skills involved. Educators can hardly expect a mastery or even an awareness of the basic . Ubicomp education should be tuned to the skills of students.

A great outcome of the seminar would be a consolidated list of small-units that could be assembled in a variety of ways, as well as write-ups and reflections on existing curricula and teaching practices:

A Pervasive course outline with resources from the ITU of Copenhagen:
https://blog.itu.dk/SPCT-F2012/
http://www.bardram.net/wp-content/papercite-data/pdf/bardram-ieaward2012.pdf

Two sources of inspiration from nearby domains: https://wot.pubpub.org/

Interaction Design: https://medium.com/ixda/education-summit/home

Visualization: http://vgl.cs.usfca.edu/pdvw/2017/http://vgl.cs.usfca.edu/pdvw/2016/

3.27 Reflections from Brygg Ullmer on ubicomp education

Brygg Ullmer (Clemson University, US)

While perhaps unsurprising, but simultaneously interesting and at times illuminating, has been discussion concern the "ubiquitous computing" term, implications of its use in the context of our respect academic institutional environments, etc. For example, several noted that in the 3-minute presentations on day 1, relatively few (perhaps 5 or fewer) of the 28 participants explicitly framed their work in terms of ubiquitous computing. Several times in day 2, participants described naming implications including (e.g.) gender engagement. As one specific example, among several sister institutions, it was noted "interactive product design" attracted a more balanced gender dynamic than "interactive product development."

For some participants, one gathered that the particular scoping and components of terms like "ubiquitous computing" are quite consequential. For many others, it seemed a wide variety of descriptive terms – often engaging particularities of individual institutions, and their orientations toward engineering, art, design, liberal arts, etc. – might often inflect both local resonance and use. A sister bridging term that several referenced is "interactive systems;" or perhaps as an extension, ISDE (interactive systems design and engineering & evaluation, per multiple communities "overloading" individual letters of acronyms). One related exercise might be an (e.g.) 5xN mapping for some of the overarching themes discussed in some breakouts on day two – e.g. people, technology, objects, methods, business – and strong, weak, or non-engagement with Ubicomp (early 90s), Ubicomp (present-future), Pervasive, ISDE, CHI, TEI, and others.

3.28 Reflections from Vicky Zeamer on ubicomp education

Vicky Zeamer (Hubspot – Cambridge, US)

Why is the training and education of students in UbiComp important for industry, not just academia? Much time at this seminar was spent debating what was even considered ubiquitous computing. Without a clear spotlight on the definition and why it's important, industry does not have a clear incentive to put resources into UbiComp education. This seminar has illustrated not only the cross-interests we have as researchers and designers related to humans and technologies, but also the lack of consensus of what it really means in practice to live and work within a world with computers being "ubiquitous."

As confused as we are by the implications for education related to UbiComp, imagine how our students and junior colleagues must feel about their own preparation and paths? I have always felt, as a student going through a pre-designed curriculum, that these programs were out of touch with the realities of real world contexts and problems. This disconnect between education and industry careers is often by design— academia is for exploring unknown frontiers and being free of the friction that real world implementations often impose. However, as professionals (either academics or practitioners) within the realm of human-technology interaction, we have a moral responsibility to prepare a majority of students to tackle the world's issues, primarily via industry.

At the end, I believe that we need to, as a group of experts, is to establish the fact that we are preparing students to be ethical leaders, builders, and users of ubiquitous computing outside of academic research environments. While this may seem like a straightforward sentiment, what I am stressing is that we must understand that a majority of students will not end up in positions like their own professors. With that in mind, how do we teach students who will shape our future world and make it a better place via UbiComp?

Participants

- Michael Beigl
KIT – Karlsruher Institut für Technologie, DE
- Andrea Bianchi
KAIST – Daejeon, KR
- Anke Brock
ENAC – Toulouse, FR
- Jessica Cauchard
Ben Gurion University – Beer Sheva, IL
- Ruzanna Chitchyan
University of Bristol, GB
- Jeremy Cooperstock
McGill University – Montreal, CA
- Ellen Yi-Luen Do
University of Colorado – Boulder, US
- Audrey Girouard
Carleton University – Ottawa, CA
- Eva Hornecker
Bauhaus-Universität Weimar, DE
- Miriam Konkel
Clemson University, US
- Andrew Kun
University of New Hampshire – Durham, US
- Sylvain Malacria
INRIA Lille, FR
- Nicolai Marquardt
University College London, GB
- Amanda McLeod
Carleton University – Ottawa, CA
- Donald McMillan
Stockholm University, SE
- Timothy Merritt
Aalborg University, DK
- Caitlin Mills
University of New Hampshire – Durham, GB
- Simon Perrault
Singapore University of Technology and Design, SG
- Thomas Pietrzak
INRIA Lille, FR
- Michal Rinott
SHENKAR – Engineering. Design. Art – Ramat-Gan, IL
- Anne Roudaut
University of Bristol, GB
- Albrecht Schmidt
LMU München, DE
- Oliver Schneider
University of Waterloo, CA
- Orit Shaer
Wellesley College, US
- Jakub Sypniewski
Universität Salzburg, AT
- Aurélien Tabard
Université de Lyon, FR
- Brygg Ullmer
Clemson University, US
- Vicky Zeamer
Hubspot – Cambridge, US

Report from Dagstuhl Seminar 19241

25 Years of the Burrows-Wheeler Transform

Edited by

Travis Gagie[1], Giovanni Manzini[2], Gonzalo Navarro[3], and Jens Stoye[4]

1 **Dalhousie University – Halifax, CA,** travis.gagie@dal.ca
2 **University of Eastern Piedmont – Alessandria, IT,** giovanni.manzini@uniupo.it
3 **University of Chile – Santiago de Chile, CL,** gnavarro@dcc.uchile.cl
4 **Universität Bielefeld, DE,** jens.stoye@uni-bielefeld.de

Abstract

Dagstuhl Seminar 19241 ("25 Years of the Burrows-Wheeler Transform") took place from June 10th to 14th, 2019, and was attended by 45 people from 13 countries and the three fields of Algorithms and Data Structures, Bioinformatics, and Combinatorics on Words. There were four talks and a panel session for each field. Feedback was generally positive and we are confident the seminar fostered interdisciplinary connections and will eventually result in noteworthy joint publications.

Seminar June 10–14, 2019 – http://www.dagstuhl.de/19241

2012 ACM Subject Classification Applied computing → Bioinformatics, Applied computing → Computational genomics, Applied computing → Molecular sequence analysis, Applied computing → Genomics, Mathematics of computing → Combinatorics on words, Theory of computation → Data compression, Theory of computation → Pattern matching, Theory of computation → Sorting and searching

Keywords and phrases Bioinformatics, Burrows-Wheeler Transform, Combinatorics on Words, Data Compression, Data Structures, Indexing, Sequence Alignment

Digital Object Identifier 10.4230/DagRep.9.6.55

1 Executive Summary

Travis Gagie
Giovanni Manzini
Gonzalo Navarro
Jens Stoye

Dagstuhl Seminar 19241 marked the 25th anniversary of the publication of the Burrows-Wheeler Transform (BWT), which has had a huge impact on the fields of data compression, combinatorics on words, compact data structures, and bioinformatics. The 10th anniversary in 2004 was marked by a workshop at the DIMACS Center at Rutgers (http://archive.dimacs.rutgers.edu/Workshops/BWT) organized by Paolo Ferragina, Giovanni and S. Muthukrishnan, and it is exciting to see how far we have come. In the past 15 years, interest in the BWT has shifted from data compression to compact data structures and bioinformatics, particularly indexing for DNA read alignment, but seven of the 33 participants of that workshop (including Giovanni) also attended this seminar. Unfortunately, Professor Gørtz fell ill at the last minute and emailed us on June 11th to say she couldn't attend, but everyone else on the

Figure 1 Gonzalo with his birthday cake (featuring a BWT).

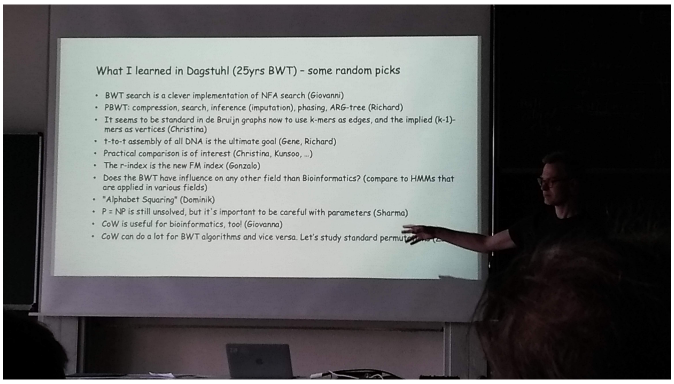

Figure 2 Jens reviewing some points raised during the seminar.

final list of invitees was present for at least some of the seminar (although not everyone made it into the photo). In total there were 45 people (listed at the end of this report) from 13 countries, including ten women, six junior researchers and two researchers from industry. By happy coincidence, the seminar started the day after Gonzalo's 50th birthday, so we were able to celebrate that as well. We thank Professor Sadakane for the photos shown in Figures 1 and 2.

The schedule, shown in Figure 3, featured an introduction, 12 talks, three panel sessions and a closing. The talks were all timely and reflected the active and dynamic research being carried out on the BWT:

- Giovanni's introduction was a more in-depth version of his invited talk from DCC '19;
- Veli Mäkinen surveyed pan-genomic indexing, including work published in *BMC Genomics* last year;
- Richard Durbin surveyed results based on the positional BWT, published in *Bioinformatics* in 2014;
- Jouni Sirén presented work included in a *Nature Biotechnology* article last year;
- Christina Boucher surveyed compact data structures for de Bruijn graphs, including work from an ISMB/ECCB 2019 paper;
- Gonzalo Navarro reviewed BWT-based indexes, including work from a SODA '18 paper;
- Sandip Sinha presented work from a STOC '19 paper;
- Dominik Kempa presented work from another STOC '19 paper;
- Sharma Thankachan presented work from an ESA '19 paper;
- Nicola Prezza presented work from a STOC '18 paper;
- Marinella Sciortino gave a version of her invited lecture for IWOCA '19 a month later;
- Giovanna Rosone presented results about two extensions of the BWT, including work from a WABI '18 paper, now published in *Algorithms for Molecular Biology*;
- Dominik Köppl presented work from a CPM '19 paper.

We later received all the abstracts but one.

	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY
07:30		BREAKFAST	BREAKFAST	BREAKFAST	BREAKFAST
09:00		INTRO	ALG TALK 1	CoW TALK 1	WORK...
09:45		BIO TALK 1	ALG TALK 2	CoW TALK 2	
10:30		BIO TALK 2	ALG TALK 3	CoW TALK 3	
11:15		BIO TALK 3	ALG TALK 4	CoW TALK 4	
12:15		LUNCH	LUNCH	LUNCH	LUNCH
13:45		BIO TALK 4			
14:00			ALG PANEL	CoW PANEL	
14:30		BIO PANEL			
15:00			WORK!	CLOSING	
15:30	CAKE	CAKE	CAKE	CAKE	
16:00	WORK?	WORK	WORK!!	WORK!!!	
18:00	DINNER (buffet)	DINNER	DINNER	DINNER	
20:00	CHEESE?	CHEESE	CHEESE	CHEESE	

INTRO	Giovanni		BIO PANEL	ALG PANEL	CoW PANEL
BIO TALK 1	Veli	(Pan-genomic) alignment	Ben	Ian	Gabriele
BIO TALK 2	Richard	PBWT	Gene	Inge (chair)	Hideo
BIO TALK 3	Jouni	GBWT	Knut	Johannes	Jackie
BIO TALK 4	Christina	de Bruijn graphs	Kunsoo	Rahul	Pawel
ALG TALK 1	Gonzalo	r-index	Paola	Roberto	Sabrina (chair)
ALG TALK 2	Sandip	Local decodability	Richard	Simon G	Tomasz
ALG TALK 3	Dominik	BWT construction	Tony (chair)		Zsuzsa
ALG TALK 4	Sharma	Wheeler graphs			
CoW TALK 1	Nicola	String attractors	Jens chairs BIO talks		
CoW TALK 2	Marinella	Combinatorial properties	Giovanni chairs ALG talks		
CoW TALK 3	Giovanna	eBWT / BWT similarity	Travis chairs CoW talks		
CoW TALK 4	Dominik	Bijective BWT			
CLOSING	Jens				

Figure 3 The original seminar schedule. Inge Li Gørtz was unable to attend and so Tatiana Starikovskaya chaired the *Algorithms and Data Structures* panel. The talks and panel on *Bioinformatics* were held on the first day and those on *Algorithms and Data Structures* on the second day to accommodate participants' schedules.

2 Table of Contents

3 Overview of Talks

Apart from the talks below, there were impromptu presentations by Jackie Daykin on "Order-based Burrows-Wheeler Transforms", by Enno Ohlebusch on "An improved encoding of genetic variation in a Burrows-Wheeler transform", by Kunsoo Park on "Comparing Pan-Genomic Indexes". Slides from all of these and introductory slides from the Combinatorics on Words panel (submitted by Sabrina Mantaci) are available on the materials page.

3.1 25 Years of Burrows-Wheeler Transform: A review

Giovanni Manzini (University of Eastern Piedmont – Alessandria, IT)

Joint work of Giovanni Manzini, Paolo Ferragina, Travis Gagie, Raffaele Giancarlo, Marinella Sciortino, Jouni Sirén
Main reference Travis Gagie, Giovanni Manzini, Jouni Sirén: "Wheeler graphs: A framework for BWT-based data structures", Theor. Comput. Sci., Vol. 698, pp. 67–78, 2017.
URL https://doi.org/10.1016/j.tcs.2017.06.016

To establish a common ground, in this introductory talk we review the main properties of the BWT with respect to data compression and text indexing.

The "Block sorting data compression algorithm" by Mike Burrows and David Wheeler [1] was based on a data transformation, now called the BWT, designed "to make redundancy in the input more accessible". While this is obvious at the intuitive level, it took ten years to formalize this notion in terms of empirical entropy. In [2] it was proven that the Burrows-Wheeler transform can be seen as a "compression booster", that is a tool for transforming any order-0 encoder into a much more effective order-k encoder, and that this result holds simultaneously for every $k > 0$.

Starting from the year 2000, several researchers [3, 5, 6] observed that, because of the relationship between the BWT and the Suffix Array, the former can be used as a sort of full text index, possibly compressed if one takes advantage of the "boosting" properties of the BWT. Over the years, these ideas have been extended to design compressed data structures to index other discrete structures such as trees, graphs, automata, alignments, and so on.

We introduced the notion of Wheeler Graph that generalizes the BWT and provides a unified view of many of these extensions [4]. We show that pattern matching problems inside many of these discrete structures can be modeled using Nondeterministic Finite Automata which have the additional property of being Wheeler Graphs. We also show that we can compactly represent and navigate Wheeler Graphs using the well-known and highly optimized rank and select operations on linear arrays. Although not every BWT-related data structure fits in our framework, we believe our unifying view can help researchers develop new BWT variants and new indexing data structures.

References

1 M. Burrows, D. Wheeler, A block-sorting lossless data compression algorithm, Tech. Rep. 124, Digital Equipment Corporation (1994).
2 P. Ferragina, R. Giancarlo, G. Manzini, M. Sciortino, Boosting textual compression in optimal linear time, J. ACM 52 (2005) 688–713.
3 P. Ferragina, G. Manzini, Opportunistic data structures with applications, in: Proc. 41st IEEE Symp. on Found. of Computer Science, 2000, pp. 390–398.

4 T. Gagie, G. Manzini, J. Sirén, Wheeler graphs: A framework for BWT-based data structures, Theor. Comput. Sci. 698 (2017) 67–78.
5 R. Grossi, J. S. Vitter, Compressed suffix arrays and suffix trees with applications to text indexing and string matching, in: Proc. of the 32nd ACM Symposium on Theory of Computing, 2000, pp. 397–406.
6 V. Mäkinen, Compact suffix array, in: Proc. of the 11th Symposium on Combinatorial Pattern Matching, Springer-Verlag LNCS n. 1848, 2000, pp. 305–319.

3.2 Scaling pan-genomic alignment using founders

Veli Mäkinen (University of Helsinki, FI)

The talk covers the PanVC framework for pan-genomic variant calling through multiple reference indexing for read alignment [1], as well as a lossy compression of multiple references into founders [2], which makes the whole framework scalable.

References

1 D. Valenzuela, Tuukka Norri, Niko Välimäki, Esa Pitkänen and Veli Mäkinen. Towards pan-genome read alignment to improve variation calling. *BMC Genomics*, Vol. 19, No. 87, 2018.
2 T. Norri, B. Cazaux, D. Kosolobov and V. Mäkinen. Linear time minimum segmentation enables scalable founder reconstruction. *Algorithms for Molecular Biology*, Vol. 14, No. 12, 2019.

3.3 Genome Graphs and BWT-based Data Structures

Jouni Sirén (University of California – Santa Cruz, US)

Joint work of Jouni Sirén, Erik Garrison, Adam M. Novak, Benedict Paten, Richard Durbin
Main reference Jouni Sirén, Erik Garrison, Adam M. Novak, Benedict Paten, Richard Durbin: "Haplotype-aware graph indexes", in Proc. of the 18th International Workshop on Algorithms in Bioinformatics, WABI 2018, August 20-22, 2018, Helsinki, Finland, LIPIcs, Vol. 113, pp. 4:1–4:13, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2018.
URL http://dx.doi.org/10.4230/LIPIcs.WABI.2018.4

A reference sequence may represent a new dataset poorly if the sequenced individual diverges substantially at some location. Mapping reads to such a reference can introduce reference bias into the subsequent analysis. Genome graphs help to avoid the bias by including genetic variation in the reference. Although each path in the graph is a potential haplotype, most paths are unlikely recombinations of true haplotypes.

In this talk, I will show how we can use BWT-based methods to index genome graphs. We transform the graph into an equivalent Wheeler graph, or approximate it with Wheeler graphs when the equivalent Wheeler graph is too large or does not exist. I will also introduce the GBWT as a fast and space-efficient way of storing large collections of haplotypes as paths over the genome graph.

3.4 BWT Meets the de Bruijn Graph: Results and Challenges

Christina Boucher (University of Florida – Gainesville, US)

The money and time needed to sequence a genome have decreased remarkably in the past decade. With this decrease has come an increase in the number and rate at which sequence data is collected for public sequencing projects. This led to the existence of GenomeTrakr, which is a large public effort to use genome sequencing for surveillance and detection of outbreaks of foodborne illnesses. This effort includes over 50,000 samples, spanning several species available through this initiative, a number that continues to rise as datasets are continually added. Unfortunately, analysis of this dataset has been limited due to its size. In this talk, I will describe our method for constructing the colored de Bruijn graph for large datasets that is based on partitioning the data into smaller datasets, building the colored de Bruijn graph using a FM-index based representation, and succinctly merging these representations to build a single graph. Finally, I will show its capability of building a colored de Bruijn graph for 16,000 strains from GenomeTrakr in a manner that allows it to be updated. Lastly, I conclude by outlining some opportunities for further study in this area.

3.5 Text Indexing with the BWT

Gonzalo Navarro (University of Chile – Santiago de Chile, CL)

Main reference Paolo Ferragina, Giovanni Manzini: "Opportunistic Data Structures with Applications", in Proc. of the 41st Annual Symposium on Foundations of Computer Science, FOCS 2000, 12-14 November 2000, Redondo Beach, California, USA, pp. 390–398, IEEE Computer Society, 2000.
URL https://doi.org/10.1109/SFCS.2000.892127

The talk covers the history and functioning of the FM-index, since its first version (Ferragina and Manzini, 2000) to the latest one (Gagie, Navarro, and Prezza, 2018) aimed at repetitive datasets.

3.6 Local Decodability of the Burrows-Wheeler Transform

Sandip Sinha (Columbia University – New York, US)

Joint work of Sandip Sinha, Omri Weinstein
Main reference Sandip Sinha, Omri Weinstein: "Local decodability of the Burrows-Wheeler transform", in Proc. of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019., pp. 744–755, ACM, 2019.
URL https://doi.org/10.1145/3313276.3316317

The Burrows-Wheeler Transform (BWT) is among the most influential discoveries in text compression and DNA storage. It is a reversible preprocessing step that rearranges an n-letter string into runs of identical characters (by exploiting context regularities), resulting in highly compressible strings, and is the basis of the `bzip` compression program. Alas, the decoding process of BWT is inherently sequential and requires $\Omega(n)$ time even to retrieve a *single* character.

We study the succinct data structure problem of locally decoding short substrings of a given text under its *compressed* BWT, i.e., with small additive redundancy r over the *Move-To-Front* (`bzip`) compression. The celebrated BWT-based FM-index (FOCS '00), as well as other related literature, yield a trade-off of $r = \tilde{O}(n/\sqrt{t})$ bits, when a single character is to be decoded in $O(t)$ time. We give a near-quadratic improvement $r = \tilde{O}(n \lg(t)/t)$. As a by-product, we obtain an *exponential* (in t) improvement on the redundancy of the FM-index for counting pattern-matches on compressed text. In the interesting regime where the text compresses to $o(n)$ (say, $n/\mathrm{polylg}(n)$) bits, these results provide an $\exp(t)$ *overall* space reduction. For the local decoding problem of BWT, we also prove an $\Omega(n/t^2)$ cell-probe lower bound for "symmetric" data structures.

We achieve our main result by designing a compressed partial-sums (Rank) data structure over BWT. The key component is a *locally-decodable* Move-to-Front (MTF) code: with only $O(1)$ extra bits per block of length $n^{\Omega(1)}$, the decoding time of a single character can be decreased from $\Omega(n)$ to $O(\lg n)$. This result is of independent interest in algorithmic information theory.

3.7 BWT Construction: History, Techniques, State of the Art, Open Problems

Dominik Kempa (University of Warwick – Coventry, GB)

Joint work of Dominik Kempa, Tomasz Kociumaka
Main reference Dominik Kempa, Tomasz Kociumaka: "String synchronizing sets: sublinear-time BWT construction and optimal LCE data structure", in Proc. of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019., pp. 756–767, ACM, 2019.
URL https://doi.org/10.1145/3313276.3316368

Burrows-Wheeler transform (BWT) is an invertible text transformation that, given a text T of length n, permutes its symbols according to the lexicographic order of suffixes of T. BWT is one of the most heavily studied algorithms in data compression with numerous applications in indexing, sequence analysis, and bioinformatics. Its construction is a bottleneck in many scenarios, and settling the complexity of this task is one of the most important unsolved problems in sequence analysis that has remained open for 25 years. In this talk, I will review the recent progress made for the problem of BWT construction [1, 2] as well as summarize the existing algorithms and outline the main challenges lying ahead.

References

1 Dominik Kempa and Tomasz Kociumaka. *String Synchronizing Sets: Sublinear-Time BWT Construction and Optimal LCE Data Structure*. In Proceedings of the 51st Annual ACM SIGACT Symposium on the Theory of Computing (STOC), 2019, ACM.
2 Dominik Kempa. *Optimal Construction of Compressed Indexes for Highly Repetitive Texts*. In Proceedings of the 30th Annual ACM SIAM Symposium on Discrete Algorithms (SODA), 2019, SIAM.

3.8 On the Hardness and Inapproximability of Recognizing Wheeler Graphs

Sharma V. Thankachan (University of Central Florida – Orlando, US)

Joint work of Daniel Gibney, Sharma V. Thankachan
Main reference Daniel Gibney, Sharma V. Thankachan: "On the Hardness and Inapproximability of Recognizing Wheeler Graphs", in Proc. of the 27th Annual European Symposium on Algorithms, ESA 2019, September 9-11, 2019, Munich/Garching, Germany., pp. 51:1–51:16, 2019.
URL http://dx.doi.org/10.4230/LIPIcs.ESA.2019.51

In recent years several compressed indexes based on variants of the Burrows-Wheeler transformation have been introduced. Some of these are used to index structures far more complex than a single string, as was originally done with the FM-index [Ferragina and Manzini, J. ACM 2005]. As such, there has been an increasing effort to better understand under which conditions such an indexing scheme is possible. This has led to the introduction of Wheeler graphs [Gagie *et al.*, Theor. Comput. Sci., 2017]. Gagie *et al.* showed that de Bruijn graphs, generalized compressed suffix arrays, and several other BWT related structures can be represented as Wheeler graphs, and that Wheeler graphs can be indexed in a way which is space efficient. Hence, being able to recognize whether a given graph is a Wheeler graph, or being able to approximate a given graph by a Wheeler graph, could have numerous applications in indexing. Here we resolve the open question of whether there exists an efficient algorithm for recognizing if a given graph is a Wheeler graph. We present:

The problem of recognizing whether a given graph $G = (V, E)$ is a Wheeler graph is NP-complete for any edge label alphabet of size $\sigma \geq 2$, even when G is a DAG. This holds even on a restricted, subset of graphs called d-NFA's for $d \geq 5$. This is in contrast to recent results demonstrating the problem can be solved in polynomial time for d-NFA's where $d \leq 2$. We also show the recognition problem can be solved in linear time for $\sigma = 1$;

There exists an $2^{e \log \sigma + O(n+e)}$ time exact algorithm where $n = |V|$ and $e = |E|$. This algorithm relies on graph isomorphism being computable in strictly sub-exponential time;

We define an optimization variant of the problem called Wheeler Graph Violation, abbreviated WGV, where the aim is to remove the minimum number of edges in order to obtain a Wheeler graph. We show WGV is APX-hard, even when G is a DAG, implying there exists a constant $C \geq 1$ for which there is no C-approximation algorithm (unless P = NP). Also, conditioned on the Unique Games Conjecture, for all $C \geq 1$, it is NP-hard to find a C-approximation;

We define the Wheeler Subgraph problem, abbreviated WS, where the aim is to find the largest subgraph which is a Wheeler Graph (the dual of the WGV). In contrast to WGV, we prove that the WS problem is in APX for $\sigma = O(1)$;

The above findings suggest that most problems under this theme are computationally difficult. However, we identify a class of graphs for which the recognition problem is polynomial time solvable, raising the open question of which parameters determine this problem's difficulty.

3.9 String Attractors

Nicola Prezza (University of Pisa, IT), Travis Gagie (Universidad Diego Portales, CL), Dominik Kempa (University of Warwick – Coventry, GB), and Gonzalo Navarro (University of Chile – Santiago de Chile, CL)

Main reference Dominik Kempa, Nicola Prezza: "At the roots of dictionary compression: string attractors", in Proc. of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018, pp. 827–840, ACM, 2018.
URL https://doi.org/10.1145/3188745.3188814

In this talk I show connections between the Burrows-Wheeler transform and popular compressors such as LZ77 and grammars. The first connection is that these compressors can be interpreted as approximation algorithms for computing a new combinatorial object: the string attractor [1]. A string attractor is a subset of the text's positions such that each distinct text substring has at least one occurrence crossing at least one element in the set. It turns out that most dictionary compressors induce string attractors whose cardinalities are bounded by their outputs' sizes, and that one can build a dictionary-compressed representation from a string attractor. It follows that these new objects allow one to prove new relations between the sizes of dictionary compressors, and to design universal compressed data structures. The second connection is through bidirectional parsings. A bidirectional parse is a generalization of LZ77 where phrases' sources are not forced to precede their destination. I will show that the BWT induces a bidirectional parse with r phrases, where r is the number of equal-letter runs in the BWT. Unlike LZ77, this parse enjoys new fascinating properties that allow one to build an optimal-time index on top of it [2].

References

1 Dominik Kempa and Nicola Prezza *At the Roots of Dictionary Compression: String Attractors*. Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, Los Angeles, CA, USA, 2018.
2 Travis Gagie, Gonzalo Navarro, and Nicola Prezza *Optimal-time Text Indexing in BWT-runs Bounded Space*. Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, Louisiana, USA, 2018.

3.10 Combinatorial Properties of BWT

Marinella Sciortino (University of Palermo, IT)

Although the BWT has been introduced in Data Compression, over the years it has found many applications in several different contexts. The outstanding versatility and efficacy of the BWT is based on some mathematical and combinatorial properties, i.e. its efficient reversibility and the "clustering effect" on the output. In this talk such properties are explored, highlighting the connections with well-known objects in Combinatorics of words, such as Lyndon words and Sturmian words. Furthermore, studying how the number of equal-letter runs varies after the BWT is applied, allows us to characterize infinite families of words based on the clustering effect produced by BWT. In such cases these characterizations are connected to still-open mathematical conjectures. Finally, some variants of the BWT are described. The first one, denoted eBWT, is defined on multisets of strings and allows to establish a bijection between the multiset of conjugacy classes of strings and all the strings on a given alphabet, with interesting theoretical and applicative implications. The second variant, denoted ABWT, uses a different order (called alternating order) to sort the cyclic rotations of a string. It is interesting to note that the ABWT preserves many combinatorial and mathematical properties of the BWT and it can be used as a compressed index in the same way as the BWT.

3.11 BWT / eBWT similarity

Giovanna Rosone (University of Pisa, IT)

Joint work of Giovanna Rosone, Veronica Guerrini, Sabrina Mantaci, Marinella Sciortino, Antonio Restivo
Main reference Sabrina Mantaci, Antonio Restivo, Giovanna Rosone, Marinella Sciortino: "A New Combinatorial Approach to Sequence Comparison", Theory Comput. Syst., Vol. 42(3), pp. 411–429, 2008.
URL https://doi.org/10.1007/s00224-007-9078-6

Sequence comparison has become a very essential tool in modern molecular biology. In fact, in biomolecular sequences high similarity usually implies significant functional or structural similarity. Traditional approaches use techniques that are based on sequence alignment able to measure character level differences. Here, we describe some similarity measures, alignment-free, based on the Burrows-Wheeler transform with several application in bioinformatics, such as the metagenomic problem.

3.12 Searching Patterns in the Bijective BWT

Dominik Köppl (Kyushu University – Fukuoka, JP)

Joint work of Hideo Bannai, Juha Kärkkäinen, Dominik Köppl, Marcin Piatkowski
Main reference Hideo Bannai, Juha Kärkkäinen, Dominik Köppl, Marcin Piatkowski: "Indexing the Bijective BWT", in Proc. of the 30th Annual Symposium on Combinatorial Pattern Matching, CPM 2019, June 18-20, 2019, Pisa, Italy., LIPIcs, Vol. 128, pp. 17:1–17:14, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2019.
URL http://dx.doi.org/10.4230/LIPIcs.CPM.2019.17

We present an index data structure for the bijective Burrows-Wheeler transform [1]. The index data structure is based on the FM index [2]. Like the FM index, it reports the suffix array interval of all pattern occurrences by means of backward searches.

References

1. Joseph Yossi Gil and David Allen Scott. A bijective string sorting transform. *ArXiv 1201.3077*, 2012. `arXiv:1201.3077`.
2. Paolo Ferragina and Giovanni Manzini. Opportunistic data structures with applications. In *Proc. FOCS*, pages 390–398, 2000.

4 Motivation (from proposal)

When it was introduced in a technical report in May 1994, no one could have foreseen the impact the Burrows-Wheeler Transform (BWT) would have far beyond the field of data compression for which it was originally intended. Of course it first made a significant impact on compression, both in theory and in practice (e.g., as the basis for bzip2). New horizons opened up in 2000 with the introduction of the FM-index, a compressed suffix array based on the BWT. Among other applications, in the next decade FM-indexes became the heart of the DNA aligners such as Bowtie, BWA and SOAP 2 that helped pave the way for the genomics revolution.

Generalizations of BWT to labelled trees, de Bruijn graphs, automata, haplotype sequences and genomic reference graphs have kept the exchange of ideas lively between researchers in algorithms and data structures, bioinformatics, combinatorics on words and information retrieval. Burrows and Wheeler's original technical report is still cited hundreds of times every year, subsequent papers are cited thousands of times, and new results about or using the BWT appear in the top conferences and journals.

By now, 25 years since its publication, probably no one person knows all the results that have been proven about the BWT, but we hope the expertise gathered together at this Dagstuhl Seminar will make progress on the following topics, among others:

FM-indexes for genomic databases: FM-indexes shine for indexing one or a few genomes, but they have not scaled well to indexing the genomic databases that have resulted from high-throughput sequencing technologies. An important problem has been that the suffix array samples used to locate occurrences of patterns must be fairly large or locating becomes very slow. Very recently, a way was discovered to greatly compress also the suffix array sample for repetitive texts, opening the door to indexing thousands of genomes. We expect this seminar will lead to a fuller understanding of this advance and how it can be applied in practice. Another challenge has been beating run-length compression of genomic databases' BWTs, by identifying additional structure.

More generalizations of the BWT: Many BWT researchers have heard of the generalizations to trees and graphs mentioned above, but it seems few except specialists in algorithms and data structures know about its recent extension to indexed parameterized and order-preserving pattern matching, few except specialists in combinatorics on words know about the alternating BWT, and few except bioinformaticians know about the positional BWT – but each of these may have applications in the other areas. Also, a partially unifying framework has recently been proposed, but there are still many open problems.

New challenges in bioinformatics: Papers on the BWT are published in many venues and no single conference brings together all the experts from algorithms and data structures, combinatorics on words, and theoretical and applied bioinformatics. This disconnect between the areas hurts us all because it prevents knowledge being shared efficiently. The BWT has recently been applied to some surprising bioinformatics problems, such as building ancestral recombination graphs and optical read mapping, and we expect other possibilities will emerge from interdisciplinary discussions.

5 Feedback

All of the 19 respondents to the survey said they would definitely attend another Dagstuhl seminar (5, from 1 to 5) and the median rating of the scientific quality was 10 out of 11. One person was neutral about the seminar inspiring new ideas for work, research or teaching; four agreed it did; and 13 agreed completely (and we do not know what the last person thought). Four people were neutral about it inspiring joint projects or publications; nine agreed it did; and six agreed completely. Two people disagreed that it led to insights from neighbouring fields; two were neutral; nine agreed; and five agreed completely. One person disagreed that it identified new research directions; four people were neutral; seven people agreed; and seven people agreed completely. The responses to the other questions were similar. The comments were generally positive, with people liking the mix of fields; the organization of the panels could have been improved, although they still offered some valuable insights and stimulated promising discussions.

6 Open Problems

Several general open problems were posed – e.g., generalizing the BWT to even more data types, merging BWTs of other data types (and, most generally, Wheeler graphs), improving the compression of BWT-based indexes for DNA readsets, applying current theory to practice – and some specific ones. Sharma Thankachan posed the problem of determining the degree of non-determinism that makes Wheeler graph recognition hard: it is currently known to take polytime when each node has at most two outgoing edges labelled with the same character, and it is NP-complete when nodes can have five outgoing edges labelled with the same character. Jackie Daykin asked about enhancing BWT performance with alphabet reordering (an issue about which Sharma now has preliminary results). Dominik Köppl posted a one-page description of an open problem ("Can we compute the Bijective BWT in linear time?") to the materials page.

Participants

- Jarno Alanko
University of Helsinki, FI
- Hideo Bannai
Kyushu University – Fukuoka, JP
- Paola Bonizzoni
University of Milan-Bicocca, IT
- Christina Boucher
University of Florida – Gainesville, US
- Marilia Braga
Universität Bielefeld, DE
- Anthony J. Cox
Illumina – Saffron Walden, GB
- Fabio Cunial
MPI – Dresden, DE
- Jackie Daykin
Aberystwyth University, GB
- Richard Durbin
University of Cambridge, GB
- Gabriele Fici
University of Palermo, IT
- Johannes Fischer
TU Dortmund, DE
- Travis Gagie
Universidad Diego Portales, CL
- Pawel Gawrychowski
University of Wroclaw, PL
- Simon Gog
eBay Inc – San Jose, US
- Roberto Grossi
University of Pisa, IT
- Wing-Kai Hon
National Tsing Hua University – Hsinchu, TW
- Tomohiro I
Kyushu Institute of Technology – Fukuoka, JP
- Juha Kärkkäinen
University of Helsinki, FI
- Dominik Kempa
University of Warwick – Coventry, GB
- Tomasz Kociumaka
Bar-Ilan University – Ramat Gan, IL
- Dominik Köppl
Kyushu University – Fukuoka, JP
- Ben Langmead
Johns Hopkins University – Baltimore, US
- Zsuzsanna Liptak
University of Verona, IT
- Veli Mäkinen
University of Helsinki, FI
- Sabrina Mantaci
University of Palermo, IT
- Giovanni Manzini
University of Eastern Piedmont – Alessandria, IT
- Ian Munro
University of Waterloo, CA
- Gene Myers
MPI – Dresden, DE
- Gonzalo Navarro
University of Chile – Santiago de Chile, CL
- Yakov Nekrich
University of Waterloo, CA
- Enno Ohlebusch
Universität Ulm, DE
- Kunsoo Park
Seoul National University, KR
- Nicola Prezza
University of Pisa, IT
- Knut Reinert
FU Berlin, DE
- Giovanna Rosone
University of Pisa, IT
- Kunihiko Sadakane
University of Tokyo, JP
- Leena Salmela
University of Helsinki, FI
- Marinella Sciortino
University of Palermo, IT
- Rahul Shah
Louisiana State University – Baton Rouge, US
- Sandip Sinha
Columbia University – New York, US
- Jouni Sirén
University of California – Santa Cruz, US
- Tatiana Starikovskaya
ENS – Paris, FR
- Jens Stoye
Universität Bielefeld, DE
- Sharma V. Thankachan
University of Central Florida – Orlando, US
- Rossano Venturini
University of Pisa, IT

Report from Dagstuhl Seminar 19261

Distributed Computing with Permissioned Blockchains and Databases

Edited by

C. Mohan[1], Beng Chin Ooi[2], and Gottfried Vossen[3]

1 **IBM Almaden Center – San Jose, US,** cmohan@us.ibm.com
2 **National University of Singapore, SG,** ooibc@comp.nus.edu.sg
3 **Universität Münster, DE,** vossen@wi.uni-muenster.de

Abstract

This seminar report contains the motivation, abstracts, and findings of Dagstuhl Seminar 19261 *Distributed Computing with Permissioned Blockchains and Databases* which took place in late June 2019. It brought together a very good mix of people from academia and industry as well as from databases and related areas for which blockchain is a current topic and who are either users or developers in that field.

Seminar June 23–28, 2019 – http://www.dagstuhl.de/19261
2012 ACM Subject Classification Information systems → Data management systems
Keywords and phrases Distributed database, blockchain, permissioned
Digital Object Identifier 10.4230/DagRep.9.6.69

1 Executive Summary

C. Mohan
Beng Chin Ooi
Gottfried Vossen

The topic of blockchains, and in particular that of permissioned blockchains, has rapidly gained interest in both the industrial and the research communities in recent years. It particularly pertains to situations where trust among several parties that are about to do business together is difficult to establish (e.g., due to organizational, financial, or timing reasons) or impossible to establish at all. A blockchain is a decentralized, distributed ledger that consists of immutable blocks containing transactions that can be accessed by any party, and that provides trust via replication over all nodes and an agreed-upon execution order of the transactions. Of particular interest are permissioned blockchains in which the associated parties are known and authenticated, yet still do not fully trust each other.

Many applications have shown interest in the concept of blockchains, since the situation just described applies to many real-world scenarios, including (global) supply chains, the Internet of Things, connected cars, manufacturing, banking, and healthcare. As a consequence, a number of players in the IT industry work on a development of the technology, and several consortia have been formed to advance the technology across industries, among them Hyperledger and R3. Moreover, a number of companies have released Blockchain-as-a-Service (BaaS) platforms, including IBM, Oracle, Amazon, Baidu, and Alibaba.

The technology has many links into the database community; however, the situation is basically like it was in the database area many years ago, when only a few systems had been released but users were on their own to figure out how to use them effectively. As the seminar has shown, many interesting issues remain to be solved, and there is a wide variety of aspects and research issues currently under investigation. Of these, the following were discussed:

- Blockchain scalability w.r.t. transaction throughput, one of the main roadblocks to business adoption
- Transaction ordering and endorsement, consensus of transaction commit
- Adjustments to the Proof of Work (PoW) consensus mechanism, other optimizations to consensus algorithms (e.g., Byzantine consensus) in the presence of transaction failures and in light of scalability
- Block validation
- Languages for smart-contract specification (e.g., Sandcastle SQL and Solidity)
- Amendments to Hyperledger Fabric, such as channels
- Cross-chain swaps using hashed timelocks
- Energy efficiency of blockchain applications

In addition, several participants reported on various working applications of blockchain technology.

2 Table of Contents

3 Motivation

A new era is emerging in the world of distributed computing with the growing popularity of blockchains. Traditionally, the Internet allows to exchange only data or information directly between two parties or users; a transaction, say, involving the purchase of an item requires a third party which can be trusted. Third parties often come in the form of a digital marketplace, a bank, or a trusted intermediary. Blockchains can eliminate third parties, since they are characterized by transparency, i.e., the blockchain content is visible to each participant, and being append-only, which is crucial for updating a blockchain. Conceptually, a blockchain is a decentralized and distributed digital ledger that consists of records representing transactions; since these records are tied to their history using hash values no existing record can be altered retroactively. The only kind of update allowed is to extend a given blockchain by additional records, which, assuming that the majority of participants does not pursue a dishonest intention, results in a stable view on transactions (which implies that not every party or node maintaining the blockchain needs to trust everybody else). The participants can verify and audit transactions, which results in a trustable workflow where participants' uncertainty regarding data security is marginal. The use of a blockchain also eliminates infinite reproducibility of digital assets; it confirms that each unit of value was transferred only once.

By storing data across its network, the blockchain eliminates the risks that come with data being held centrally, yet opens up for an application of distributed technology that was previously developed in other contexts. Blockchains come in two flavors: An open, permissionless, or public, blockchain network does not require any guarding against bad actors, and no access control is needed; anybody can join and leave. Hence applications can be added to the network without the approval or trust of others, using the blockchain as a transport layer. Permissioned (private) blockchains are emerging as open source protocols where openness and collaboration are encouraged among authenticated participants. They can hence restrict who can participate in the consensus processes as well as who can transact.

From a database point of view, a blockchain can be considered as a log of ordered transactions, since nodes keep replicas of the data and agree on an execution order of the transactions. A key property is the assumption that nodes behave in an arbitrary or Byzantine fashion. By being able to tolerate Byzantine failure by design, a blockchain offers stronger security than a database system. Although enterprise-grade database systems support applications like security trading and settlement, asset and finance management, or banking and insurance, blockchain technology has the potential to disrupt the status quo since they incur lower costs of infrastructure and human labor. In particular the immutability and transparency of a blockchain reduce human error as well as the need for manual intervention due to conflicting data.

While there is currently no standard in the blockchain space, all the ongoing efforts involve some combination of database, transaction, encryption, consensus and other distributed systems technologies. Some of the application areas in which blockchain pilots are being carried out are: smart contracts, supply chain management, know your customer, derivatives processing and provenance management. The seminar has surveyed some of the ongoing blockchain projects with respect to their architectures in general and their approaches to some specific technical areas. Its focus has been on how the functionality of traditional and modern data stores are being utilized or not utilized in different blockchain projects.

4 Topic Areas Discussed

This section lists the abstracts of talks given, ordered by the topic areas to which they belong.

Goals and Current State-of-the-Art of Blockchain Technology and Systems

4.1 State of Public and Private Blockchains: Myths and Reality

C. Mohan (IBM Almaden Center – San Jose, US)

It has been a decade since the concept of blockchain was invented as the underlying core data structure of the permissionless or public Bitcoin cryptocurrency network. Since then, several cryptocurrencies, tokens and ICOs have emerged. After much speculation and hype, a significant number of them have become problematic or worthless! The public blockchain system Ethereum emerged by generalizing the use of blockchains to manage any kind of asset, be it physical or purely digital, with the introduction of Smart Contracts. Over the years, numerous myths have developed with respect to the purported utility and the need for public blockchains. The adoption and adaptation of blockchains and smart contracts for use in the permissioned or private environments is what I consider to be useful and of practical consequence. Hence, the technical aspects of only private blockchains will be the focus of my talk. Along the way, I will bust many myths associated with public blockchains. I will also compare traditional database techniques with blockchain systems' features and identify desirable future research topics.

References

1 C. Mohan. State of Public and Private Blockchains: Myths and Reality. Proc. ACM SIGMOD International Conference on Management of Data, Amsterdam, July 2019. `http://bit.ly/sigBcP`
2 C. Mohan. Permissioned/Private Blockchains and Databases. `http://bit.ly/CMbcDB`

4.2 Introduction to Hyperledger

Hart Montgomery (Fujitsu Labs of America Inc. – Sunnyvale, US)

In this talk, I introduced Hyperledger, explained its structure and governance, and showed how to participate and contribute. Hyperledger is a "greenhouse" under the Linux foundation for permissioned blockchains. It is currently the largest and most popular permissioned blockchain project.

4.3 Usages of Blockchain Technologies for Data Stores

Bernhard Mitschang (Universität Stuttgart, DE)

Currently, blockchain technologies are seen as the foundation of a new business world: it will change the way the economy runs and thus will change the way we act and work, all triggered by means of some new ways to organize the relevant application data, e.g., in the areas of supply chain, health, and event storage. Blockchain technologies and systems are still in constant change and development. Hence, it is difficult to exactly define its ingredients and properties of the underlying technologies.

After having identified these technologies and associated characteristics (like transparency, provenance, fault tolerance, immutability, and authenticity), it is important to isolate and separate them into components that are subsequently used to enhance existing data stores as needed. Important questions that arise in this context are:

- How do certain Blockchain technologies and applications match?
- How to identify and separate Blockchain technologies?
- How to "append/integrate" Blockchain technologies to/with existing data stores?

Cryptography and Blockchain

4.4 Privacy, Confidentiality, Cryptography, and Security Modelling in Permissioned Blockchains

Hart Montgomery (Fujitsu Labs of America Inc. – Sunnyvale, US)

Achieving desired privacy and confidentiality properties on a blockchain can be quite difficult. This is especially true on permissioned blockchains, where it may be more difficult to hide or anonymize identities than on a public blockchain. In this talk, I explained some of the challenges that are commonly faced when attempting to achieve privacy and confidentiality on permissioned blockchains and how to go about using existing tools to achieve these properties.

One of the most important things when designing secure permissioned blockchains is the need for security modelling. Many people today pick cryptographic tools, apply them to blockchains, and then try to analyze the security properties that they get (if they even do that). This isn't a good idea for many reasons, but, in particular, it often means that blockchains do not provide the security guarantees that people want on a blockchain. For instance, even if transactions on the blockchain are encrypted or hashed, it could be the case that traffic analysis completely reveals the participants in a transaction or even information about the contents of transactions [5]. Intuitively, one might expect encryption might prevent such leakage, but it turns out that other "side channel" information on the blockchain nullifies some of the security properties of encryption.

Another very important discussion point was the notion of privacy and anonymity, and the fact that the two aren't equivalent. Many blockchain practitioners (both of the public blockchain and permissioned blockchain kind) frequently equate the two, and many disastrous consequences can happen from this. As an example, I showed my credit card history from a week last year, which, if extended further, would easily deanonymize me. Solving this issue on a blockchain is a difficult task, and blockchain builders may not want to provide perfect privacy to their users (in some cases, functionality even demands imperfect privacy, like when KYC regulations apply). However, blockchain implementers certainly need to take into account privacy (and anonymity) into their security models when building blockchains.

We next discussed cryptographic tools that can be useful for obtaining various privacy and confidentiality properties in blockchain. The first topic was threshold signatures, which allow a cryptographic signing key to be split into n different shares such that any t out of the n shares are required to create a valid signature (and any $t - 1$ shares cannot "do anything") [1]. We went through the security game of threshold signatures in detail, which illustrated how one should look at a security game for a blockchain. We also briefly defined functional encryption [3] and explained why it would be very useful for a blockchain.

The next technique discussed was "channels." Channels, developed in Hyperledger Fabric [2], are a tool intended to enable private transactions on blockchains. The idea is that each channel acts as a private "sub-blockchain" for a limited number of participants, and that people without permission do not have visibility into what is going on inside the channel. While channels are a useful tool, they have not quite reached their full potential, so if they are to be used to achieve strong privacy requirements, more development on top of them is generally required.

The final technique we discussed was trusted execution environments (TEEs). Although they have been much maligned recently in terms of their security properties [4], TEEs such as Intel's SGX offer many potential benefits for secure and private blockchains. It is possible to essentially run blockchain nodes inside TEEs such that (assuming the TEEs are secure), even the blockchain node hosts cannot see what the blockchain is doing. TEEs could potentially give us very strong privacy and confidentiality options on blockchain if they can, in fact, be built securely.

The talk ended with many questions. Overall, the goal was to expose what was mostly an audience of researchers focused on databases to some of the privacy, security, and confidentiality challenges present on blockchain today.

References

1. Boldyreva, Alexandra: Threshold signatures, multisignatures and blind signatures based on the gap-Diffie-Hellman-group signature scheme; International Workshop on Public Key Cryptography. Springer, Berlin, Heidelberg, 2003.
2. Androulaki, Elli, et al: Hyperledger fabric: a distributed operating system for permissioned blockchains; Proceedings of the Thirteenth EuroSys Conference. ACM, 2018.
3. Boneh, Dan, Amit Sahai, and Brent Waters: Functional encryption: Definitions and challenges; Theory of Cryptography Conference. Springer, Berlin, Heidelberg, 2011.
4. Kocher, Paul, et al.: Spectre attacks: Exploiting speculative execution; arXiv preprint arXiv:1801.01203 (2018).
5. Murdoch, Steven J., and George Danezis: Low-cost traffic analysis of Tor; 2005 IEEE Symposium on Security and Privacy (S & P'05). IEEE, 2005.

4.5 A Hybrid Blockchain Architecture for Enhancing Privacy and Accountability

Murat Kantarcioglu (University of Texas – Dallas, US)

Joint work of Murat Kantarcioglu, Harsh Desai, Lalana Kagal

Unfortunately, existing public blockchains and smart contracts deployed on them may disclose sensitive information. Although there is some ongoing work that leverage advanced cryptography to address some of these sensitive information leakage issues, they require significant changes to existing and popular blockchains such as Ethereum and are usually computationally expensive. On the other hand, private blockchains have been proposed to allow more efficient and privacy-preserving data sharing among pre-approved group of nodes/participants. Although private blockchains address some of the privacy challenges by allowing sensitive data to be only seen by the select group of participants, they do not allow public accountability of transactions since transactions are approved by known set of users, and cannot be accessed publicly. Given these observations, one natural question that arise is, can we leverage both public and private blockchain infrastructures to enable efficient, privacy enhancing and accountable applications ? In this talk, we try to address this challenge in the context of digital auctions.

Mainly, we discuss a novel hybrid blockchain architecture [1] that combines private and public blockchains to allow sensitive bids to be opened on a private blockchain so that only the auctioneer can learn the bids, and no one else. At the same time, we leverage public blockchains to make the auction winner announcement, and payments accountable [2]. Furthermore, using smart contracts deployed on public blockchain, we show how to incentivize truthful behavior among the auction participants. Our extensive empirical results show that this architecture is more efficient in terms of run time and monetary cost compared to pure public blockchain based auction implementations.

References

1 Harsh Bimal Desai, Murat Kantarcioglu, and Lalana Kagal. A hybrid blockchain architecture for privacy-enabled and accountable auctions. In *The 2019 IEEE International Conference on Blockchain (Blockchain-2019)*, 2019.
2 Aravind Ramachandran and Murat Kantarcioglu. Smartprovenance: A distributed, blockchain based dataprovenance system. In *Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy, CODASPY 2018, Tempe, AZ, USA, March 19-21, 2018*, pages 35–42, 2018.

Data Models

4.6 Sandcastle: a SQL Ethereum Smart Contract Language

Shahan Khatchadourian (ConsenSys – Toronto, CA)

Enterprises rely on data management frameworks in order to serve their customer. However, enterprises face challenges when integrating blockchains with existing enterprise stacks in a way that makes it easy to query, understand, and transact across systems. Challenges arise due to the complex composition of database and blockchain paradigms. As well, developers face the challenge of writing smart contracts in a low-level language like Solidity, with a need to understand concepts like decentralization, smart contracts, consensus and identity. This leads to developers building adhoc, incongruent solutions at application or protocol layers.

We propose Sandcastle, a SQL Ethereum smart contract language that integrates enterprise data management. Sandcastle works on all Ethereum blockchains (without modification or configuration), including public, private, permissioned, and permissionless networks. We showcase Sandcastle's relational features such as aggregation, triggers, functions, indexes, and row-based semantics in finance, electronic medical records, and governance use cases. We give architectural details, including the translation of Sandcastle SQL to Solidity. The Sandcastle roadmap includes optimization in performance, cost, and security. Sandcastle aims to help traditional enterprises build scalable, data-oriented blockchain platforms that span databases, Ethereum 1.0, and Ethereum 2.0 stacks and networks.

4.7 Blockchained Event Store

Dennis Przytarski (Universität Stuttgart, DE)

Consider different scenarios such as transportation/trucking and supply chain integrity. All these scenarios have one commonality: different parties generate events that need to be shared among themselves in an immutable and tamper-resistant manner. Because the stored events are used for forecasts, reports, or further process optimizations, powerful querying capabilities both on current and historical states are needed.

In general, the blockchain technology is suitable for these scenarios because it offers data immutability and tamper-resistance. For typical blockchain systems that assume transferable assets (i.e., transfer ownership of an object from one person to another person), the key-value data model and a simple query engine to answer queries such as "is a particular transaction included in a particular block" are sufficient enough.

As soon as either the data model or the query requirements increase, this basic blockchain setup does not suffice anymore. Instead, powerful query and data model capabilities are needed with immutability and tamper-resistance guaranteed. Therefore, I propose using triples (entity, attribute, value) as simple but powerful and flexible data model.

I am currently working on embedding the triple data model into a blockchain architecture with a powerful query engine. This will lead to an immutable, shared, tamper-resistant, and queryable data store for events. I am currently facing the following challenges: Data Model:

- How are changes to the data model done?
- How is the data in the triple data model (tamper-resistant) stored?

Query Language and Processing:

- How could the query language look like when there is a history to query?
- How to process a query on immutable data in the triple data model?

Consensus Protocols and Blockchain

4.8 ExpoDB Fabric: Efficient Transaction Processing in Byzantine Fault Tolerant Environments

Mohammad Sadoghi Hamedani (University of California – Davis, US)

The byzantine fault-tolerance model, studied in ExpoDB Fabric [6, 4, 5, 3, 2, 1], captures a wide-range of failures–common in real-world scenarios–such as ones due to malicious attacks and arbitrary software/hardware errors. We propose Blockplane [2], a middleware that enables making existing benign systems tolerate byzantine failures. This is done by making the existing system use Blockplane for durability and as a communication infrastructure. Blockplane proposes the following: (1) A middleware and communication infrastructure to make an entire benign protocol byzantine fault-tolerant, (2) Ahierarchical locality-aware design to minimize the number of wide-area messages, (3) A separation of fault-tolerance concerns to enable designs with higher performance.

We further investigate a protocol-agnostic approach to improve the design of primary-backup consensus protocols. At the core of our approach is a novel wait-free design of running several instances of the underlying consensus protocol in parallel [3]. To yield a high-performance parallelized design, we present coordination-free techniques to order operations across parallel instances, deal with instance failures, and assign clients to specific instances. Consequently, the design we present is able to reduce the load on individual instances and primaries, while also reducing the adverse effects of any malicious replicas. Our design is fine-tuned such that the instances coordinated by non-faulty replicas are wait-free: they can continuously make consensus decisions, independent of the behavior of any other instances.

We further develop DeltaBFT, a novel consensus protocol in which all algorithms necessary for normal-case operation only require linear communication costs, even if replicas fail [5]. At the center of our design is the delayed-replication algorithm, an algorithm we propose to reliably broadcast consensus decisions made by some non-faulty replicas to all replicas without any coordination and with low communication cost for all replicas involved. The delayed-replication algorithm is supported by our partial consensus algorithm, which uses threshold signatures to efficiently make consensus decisions.

The development of fault-tolerant distributed systems that can tolerate Byzantine behavior has traditionally been focused on consensus protocols, which support fully-replicated designs. For the development of more sophisticated high-performance Byzantine distributed systems, more specialized fault-tolerant communication primitives are necessary. As a result, we identify an essential communication primitive and study it in depth. In specifics, we formalize the cluster-sending problem [4], the problem of sending a message from one Byzantine cluster to another Byzantine cluster in a reliable manner. We not only formalize this fundamental problem, but also establish lower bounds on the complexity of this problem under crash failures and Byzantine failures. Furthermore, we develop practical cluster-sending protocols that meet these lower bounds and, hence, have optimal complexity. As such, our work provides a strong foundation for the further exploration of novel designs that address challenges encountered in fault-tolerant distributed systems.

References

1 S. Gupta and M. Sadoghi. *Blockchain Transaction Processing.* 2018.

2 F. Nawab and M. Sadoghi. Blockplane: A global-scale byzantizing middleware. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pages 124–135, April 2019.

3 S. Gupta, J. Hellings, and M. Sadoghi. Brief Announcement: revisiting consensus protocols through wait-free parallelization. In *33rd International Symposium on Distributed Computing (DISC 2019).*

4 J. Hellings, and M. Sadoghi. The fault-tolerant cluster-sending problem. In *33rd International Symposium on Distributed Computing (DISC 2019).*

5 J. Hellings, and M. Sadoghi. Brief Announcement: Byzantine fault-tolerant consensus with almost linear communication complexity. In *CoRR. 2019.*

6 M. Sadoghi, S. Blanas. Transaction Processing on Modern Hardware. In *Synthesis Lectures on Data Management, Morgan & Claypool Publishers 2019.*

4.9 CUB, a Consensus Unit-based Storage Scheme for Blockchain System

Lei Chen (HKUST – Kowloon, HK)

Joint work of Zihuan Xu, and Siyuan Han
Main reference Zihuan Xu, Siyuan Han, Lei Chen: "CUB, a Consensus Unit-Based Storage Scheme for Blockchain System", in Proc. of the 34th IEEE International Conference on Data Engineering, ICDE 2018, Paris, France, April 16-19, 2018, pp. 173–184, IEEE Computer Society, 2018.
URL https://doi.org/10.1109/ICDE.2018.00025

Recently, Blockchain becomes a hot research topic due to the success of Blockchain in many applications, such as cryptocurrency, smart contract, digital assets, distributed cloud storage and so on. The power of Blockchain is that it can achieve the consensus of an ordered set of transactions among nodes which do not trust each other, even with the existence of malicious nodes. However, compared to traditional databases, the current Blockchain technology still cannot handle a massive number of transactions, which is caused by many factors, such as the consensus protocol, structure of the blocks and storage challenge. Among them, the high storage requirement is a key factor that prevents the wide usage of Blockchain on various devices such as mobile phones or low-end PCs.

In this talk, I will discuss a novel concept called Consensus Unit (CU), which organizes different nodes into one unit and lets them to store at least one copy of Blockchain data in the system together. Based on this idea, the Blocks Assignment Optimization (BAO) problem is defined to determine the optimal assignment of blocks such that the storage space is fully used and the query cost is minimized. The problem is NP-hard. Thus, three efficient heuristic algorithms are presented to solve the static assignment problem. Furthermore, the dynamic scenarios are discussed when new blocks arrive or nodes join or depart from the CU. At the end of this talk, I will highlight some future research directions on Block chain systems.

4.10 Beyond Consensus in Permissioned Ledgers: Experiences in Using BFT Replication on DLTs

Alysson Neves Bessani (University of Lisbon, PT)

Permissioned Blockchains such as Hyperledger Fabric and R3 Corda rely on modular consensus-as-a-service components for ordering transactions. In this talk I explained how these components can be implemented using "traditional" consensus protocols (such as PBFT) and argued that such protocols are only the first step in building a robust and efficient service for these blockchains. I also showed how the required features were implemented in BFT-SMART, a replication library used for implementing Byzantine-resilient consensus-as-a-service components for both Fabric and Corda.

4.11 Red Belly Blockchain: Byzantine Consensus Is Back but Is It the Same?

Vincent Gramoli (The University of Sydney, AU)

Byzantine Consensus was proposed in the early 80's for multiple machines to reach agreement on a unique value. A practical solution, called PBFT, used a leader for implementing a network file system in a local area network in 1999. Today, with the advent of Blockchain, various companies are now trying to avoid double spending by having a large number of machines reach an agreement upon a block at any given index of the blockchain. Most companies take off-the-shelf leader-based Byzantine consensus protocols, inspired by PBFT, to solve this old consensus problem.

The issue is that the Blockchain Consensus is different from this classic Byzantine Consensus problem because the number of machines that should agree is large. Our recent design of the Democratic Practical Byzantine Fault Tolerant (DBFT) consensus algorithm solves a variant of the Byzantine Consensus problem that allows to scale by leveraging the cryptographic primitive of the blockchain to decide whether a proposal is valid. It contrasts with off-the-shelf solutions in that it is fully decentralised and does not rely on a leader to avoid bottlenecks.

The blockchain we built using DBFT, called Red Belly Blockchain, is a community blockchain whose set of consensus participants changes over time. Red Belly Blockchain uses the ECDSA public-key cryptosystem, it verifies all cryptographically signed transactions in an efficient way to avoid CPU wastage, it involves all participants to collaboratively solves this Blockchain Consensus instead of relying on a leader bottleneck. It resolves conflicts between transactions, never forks and is provably starvation-free. Our experiments show that Red Belly scales to 1000 replicas spread across 4 different continents with an average latency of 3 seconds and its peak throughput exceeds 660,000 TPS.

Performance

4.12 Enhancing Performance, Scalability, and Confidentiality of Permissioned Blockchains

Divyakant Agrawal (University of California – Santa Barbara, US)

Joint work of Divyakant Agrawal, Mohammad Java Amiri, Amr El Abbadi

Blockchains have unique features, such as transparency, provenance, fault tolerance, and authenticity, which appeal to a wide range of distributed applications, e.g., supply chain management and healthcare. However blockchain systems suffer from *performance*, *scalability*, and *confidentiality* limitations.

Existing blockchains mostly utilize an order-execute architecture where nodes agree on a total order of the blocks of transactions using a consensus protocol and then the transactions are executed in the same order on all nodes sequentially. The sequential execution of transactions on all nodes, however, reduces the blockchain performance in terms of throughput and latency. While Hyperledger Fabric increases the performance of blockchains by switching the order of the execution and ordering phases and executing the transactions in parallel, it performs poorly on workloads with high-contention, i.e., many *conflicting transactions* in a block, due to its high abort rate. To address this problem, we introduce a permissioned blockchain system *ParBlockchain* [1]. ParBlockchain is mainly introduced to support distributed applications processing workloads with *some degree of contention.* ParBlockchain consists of *orderers* and *agent* nodes. Orderers establish agreement on the order of the transactions of different applications, construct the blocks of transactions, and generate a *dependency graph* for the transactions within a block. A dependency graph enables higher concurrency by allowing the parallel execution of non-conflicting transactions. The agents of each application are then responsible for executing the transactions of that application following the generated dependency graph.

Scalability is one of the main roadblocks to business adoption of blockchain systems. Despite recent intensive research on using sharding techniques to enhance the scalability of blockchain systems, existing solutions do not efficiently address cross-shard transactions. We introduce a permissioned blockchain system, *SharPer* [2], that enhances the scalability of blockchain systems by clustering (partitioning) the nodes and assigning different data shards to different clusters. SharPer supports both intra-shard and cross-shard transactions and processes intra-shard transactions of different clusters as well as cross-shard transactions with no overlapping clusters simultaneously. In SharPer, the blockchain ledger is formed as a directed acyclic graph where each cluster maintains *only* a view of the ledger. SharPer also incorporates a protocol to establish consensus on the order of cross-shard transactions among *only* the involved clusters.

Many distributed applications need to collaborate with each other following service level agreements to provide different services. Distributed applications are often designed and implemented in different blockchain systems. In this case, inter-application collaboration could be performed as an atomic cross-chain swap, however, such an operation could negatively affect the performance of the blockchain. Furthermore, while collaboration between applications, e.g., cross-application transactions, should be *visible* to all applications, the internal data of each application, e.g, internal transactions, might be *confidential.* To support both internal and cross-application transactions of collaborating distributed applications, a permissioned blockchain system, *CAPER* [3], is introduced. In CAPER, the blockchain ledger is formed as a *directed acyclic graph* where each application accesses and maintains only its own *view* of the ledger including its internal and all cross-application transactions. CAPER also introduces three consensus protocols to globally order cross-application transactions between applications.

References

1 Mohammad Javad Amiri, Divyakant Agrawal, and Amr El Abbadi. *ParBlockchain: Leveraging Transaction Parallelism in Permissioned Blockchain Systems*. In Proceedings of the 39th IEEE International Conference on Distributed Computing Systems (ICDCS'19), pp. 1337-1347, Dallas, 2019.

2 Mohammad Javad Amiri, Divyakant Agrawal, and Amr El Abbadi. *On Sharding Permissioned Blockchains*. In Proceedings of the Second IEEE International Conference on Blockchain, pp. 282-285, Atlanta, 2019.

3 Mohammad Javad Amiri, Divyakant Agrawal, and Amr El Abbadi. *CAPER: A Cross-Application Permissioned Blockchain*. The 45th International Conference on Very Large Data Bases (VLDB'19), PVLDB 12(11), pp. 1385-1398, Los Angeles, 2019.

4.13 Hyperledger Fabric's Read-Set Conflicts and Conflict-Free Replicated Datatypes

Hans-Arno Jacobsen (TUM, DE & Univ. Toronto, CA) and Pezhman Nasirifard (TU München, DE)

Permissioned blockchains such as Hyperledger Fabric provide a robust ecosystem for developing enterprise and production-grade decentralized applications. However, the additional latency between the execution and committing the transactions, due to Fabric's adapted transaction lifecycle, is a potential scalability bottleneck. This latency can increase the probability of the occurrence of conflicting transactions, leading to the failure of a high number of transactions, which increases the application development complexity and decreases the Fabric's throughput and availability. We study an approach for integrating Conflict-Free Replicated Datatypes (CRDTs) to Hyperledger Fabric, to understand how CRDTs can improve the Fabric's availability and scalability. CRDTs are abstract data types that can resolve conflicts automatically in the presence of concurrent updates without coordination.

4.14 FastFabric: Scaling Hyperledger Fabric to 20,000 Transactions per Second

Srinivasan Keshav (University of Waterloo, CA)

Joint work of Christian Gorenflo, Stephen Lee, Lukasz Golab, Srinivasan Keshav
Main reference Christian Gorenflo, Stephen Lee, Lukasz Golab, Srinivasan Keshav: "FastFabric: Scaling Hyperledger Fabric to 20, 000 Transactions per Second", CoRR, Vol. abs/1901.00910, 2019.
URL https://arxiv.org/abs/1901.00910

Blockchain technologies are expected to make a significant impact on a variety of industries. However, one issue holding them back is their limited transaction throughput, especially compared to modern enterprise database systems. We have re-architected Hyperledger

Fabric to increase transaction throughput to 20,000 transactions per second. We focus on performance bottlenecks beyond consensus, proposing architectural changes that reduce computation and I/O overhead during transaction ordering and validation. Notably, our optimizations are fully plug-and-play and do not require any changes to Hyperledger Fabric.

4.15 Blockchains and Distributed Databases, a Twin Study

PingCheng Ruan (National University of Singapore, SG) and Beng Chin Ooi (National University of Singapore, SG)

Since the rise of Bitcoin, the public are stirring tremendous hype on its underlying blockchain technology. Over the years, the scope of blockchains has long been limited to cryptocurrency. Since the introduction of the smart contract, blockchains start to support general transactional workload, as RDBMs do. Further considering about their distributed nature, a proliferation of literature start to draw a parallel between blockchains and distributed databases. However, they mainly focus on their distinct properties to applications, but fail to identify their common technical aspects.

In this paper, we perform a joint study on blockchains and distributed databases and show that both are a twin of distributed transactional systems, with the former focusing on security while the latter on efficiency. On this common basis, we abstract out four technical aspects, replication, sharding, transaction management and storage, to lay out our comparison. Throughout, we show how the security–efficiency trade-off implicate their design goals and architectural choices. Next, we conduct an extensive performance study on two blockchains, QUORUM and FABRIC, with three distributed databases, Cockroach DB, TiDB and etcd. Our results indicate that even though the performance of blockchain is still far behind distributed databases, blockchains may still outperform them in some specific transactional workloads.

4.16 Optical Chips

Yong Tang (Univ. of Electronic Science & Technology – Chengdu, CN)

The current hardware solutions for PoW are CMOS based ASIC chips, which are slow and energy-consuming. Considering that the CMOS has met the limits of physics, it's hard to further improve speed. Moreover, it's hard to avoid energy wasting. In this talk, I introduce a design of an optical chip for PoW computations. Most of the PoW can be implemented using optical components. It's promising and attractive to do PoW with optical chips which might save energy and enjoy high speed. The possibility of doing PoW with innovative solutions such as optical chips might lead to reconsiderations of PoW and the design of cryptocurrencies.

4.17 Minimizing Transaction Failures in Permissioned Blockchains

Jeeta Ann Chacko (TU München, DE) and Hans-Arno Jacobsen (TUM, DE & Univ. Toronto, CA)

Permissioned blockchains have generally two models, namely the order-execute model and the execute-order-validate model. The order-execute model orders the incoming transactions based on a consensus algorithm and then executes the transactions on every peer in the blockchain network. Quorom, Tendermint and Ripple are examples of permissioned blockchains that follow this model. Hyperledger Fabric, on the other hand, uses the execute-order-validate model. Here the transactions are initially executed on specific peers known as endorsers which endorse these transactions. Endorsed transactions are then ordered based on a distributed consensus algorithm. The ordered transactions are then validated and committed by every peer. Both the permissioned blockchain models can be compared to database systems in certain aspects. The distributed consensus algorithms used to order the transactions are used also in replicated databases to reach consensus. Also, the order-execute-validate model is similar to the optimistic concurrency control model which has been used in various database systems. Given these parallels, it is a fruitful research direction to integrate existing database optimization strategies to improve permissioned blockchains. Our research goal is to minimize the transaction failures in permissioned blockchains. We are currently focusing on the Hyperledger Fabric implementation. The main types of of transaction failures in Hyperledger Fabric is MVCC read conflicts (inter block and intra block), phantom reads and endorsement failures. The first research area we are exploring is to use transaction reordering to reduce the number of transaction abortions. Transaction reordering is a well-known database optimization technique for databases that use optimistic concurrency control. We first create a conflict graph to find the transaction dependencies, then the minimum feedback vertex set is detected and finally the transactions are topologically sorted to minimize transaction abortion. A similar approach has been successfully used in [1] with good results. Our work differs from [1] in one aspect. We used an exact algorithm that has an exponential complexity to detect the minimum feedback vertex set. This resulted in high latency during the ordering phase resulting in more inter block MVCC read conflicts. Therefore, we were not able to show total reduction of transaction failures even though the intra block MVCC read conflicts were reduced. Currently our focus in on early commit of independent transactions and immediate re-endorsement of dependent transactions to counter the latency in the ordering phase.

References

1 Ankur Sharma, Felix Martin Schuhknecht, Divya Agrawal, and Jens Dittrich. *Blurring the Lines between Blockchains and Database Systems: the Case of Hyperledger Fabric*. In Proceedings of the 2019 International Conference on Management of Data (SIGMOD '19). ACM, New York, NY, USA, 105-122.

4.18 Blockchain Goes Green? A Time Energy Performance Study of Blockchain on Low Power Systems

Dumitrel Loghin (National University of Singapore, SG)

Permissionless blockchains are well-known to be energy inefficient, mainly because of their compute-intensive Proof-of-Work consensus protocols. On the other hand, the energy profile of permissioned blockchains is less studied. With the increasing performance of low-power, wimpy devices based on ARM or x86/64 CPUs, our goal is to analyze their time-energy performance when running blockchain applications, in comparison with traditional, brawny servers. In this work, we select three wimpy systems with power profiles in the range 5-25W, namely, (i) an Intel NUC with Intel Core i3 CPU, (ii) a Jetson TX2 with 64-bit ARM CPU and (iii) a Raspberry Pi 3 with 32-bit ARM software stack. We run BLOCKBENCH on three blockchains, namely, Hyperledger Fabric v0.6, Ethereum and Parity, in a private, permissioned setup. We show that low-end wimpy nodes, such as Raspberry Pi 3, are struggling to run full-fledged blockchains due to their small memory size and low I/O bandwidth. However, higher-performance wimpy nodes, such as Jetson TX2, achieve around 80% and 30% of the throughput of Xeon servers for Parity and Hyperledger, respectively, while using 18x and 23x less energy.

Applications

4.19 Blockchain and New Economies

Feida Zhu (SMU – Singapore, SG)

URL https://symphonyprotocol.com/

Despite its most successful and well-known application for cryptocurrencies, it is our belief that the true power of blockchain technology is to unleash the great potential of a whole class of virtual assets, whose value are long known but not yet well established. Such assets include data, influence, social network, credit, to name a few. As an example, I will demonstrate in this talk how blockchain technology can be used to establish individual data as an emerging asset class to solve the bottleneck in today's data-driven economy. We will examine the key issues we face today from both the perspectives of the businesses and the individual users, and explore how blockchain-based platform could provide both the "trust" and "incentive" necessary to foster a self-growing data ecosystem. We introduce "Symphony Protocol", which is a blockchain-based protocol to create an ecosystem that unlocks personal data for democratized and personalized intelligence, with privacy by design.

4.20 Blockchain-based Cross-Site Genomic Dataset Access Audit

Li Xiong (Emory University – Atlanta, US)

Genomic data have been collected by different institutions and companies and need to be shared for broader use. In a cross-site genomic data sharing system, a secure and transparent access control audit module plays an essential role in ensuring the accountability. The goal of the iDASH competition 2018 first track is to develop blockchain-based ledgering solutions to log and query the user activities of accessing genomic datasets across multiple sites. We designed a Multichain-based log system which can provide a light-weight and widely compatible module for existing blockchain platforms. The submitted solution won the third place of the competition. Our method introduces an on-chain indexing data structure which can be easily adapted to any blockchains that use key-value database as their local storage.

4.21 Leveraging Decentralized, Secure and Governed Exchange of Confidential Information with Permissioned Blockchain

Gabriela Ruberg (Banco Central do Brasil – Rio de Janeiro, BR)

Joint work of Gabriela Ruberg, Marcus Vinicius Cursino, Jose Deodoro, Rafael Sarres, Aristides Cavalcante-Neto

The impressive popularity of blockchain applications, such as Bitcoin, has fostered the emergence of a variety of software tools to develop decentralized P2P systems. This has opened up the way for several new possibilities to explore blockchain technology beyond cryptocurrencies and financial services. In particular, permissioned blockchain networks (that is, when participants are identified and previously authorized) allow benefiting from relevant blockchain properties, especially tamper-proof data and non-repudiation, with better performance.

Sharing confidential data among autonomous entities in a secure and governed environment remains a challenge that can benefit from this new blockchain perspective. In practice, canonical solutions involving either centralized databases or traditional information integration are not sufficient nor sustainable. They usually require significant up-front efforts and cannot easily support updates with new datasets and views. Also, they present long time-to-data (namely, the time for new information to be available), require frequent (and expensive!) data transfers and lack trustful data governance. In many cases, choosing trusted third parties is not trivial. Moreover, recent regulation on data protection has further highlighted the disadvantages of siloed-data solutions.

This problem is relevant, for instance, in the context of public agencies and regulators, which need to frequently exchange protected data in order to perform due diligence processes and to provide integrated public services.

To tackle these issues, at the Central Bank of Brazil we developed a blockchain platform, called PIER, to enable entities to share, integrate and exchange sensitive data in a flexible, secure and governed environment. The PIER platform runs a permissioned blockchain network where participant nodes can easily discover and publish datasets from off-chain data sources, and then share metadata on the available datasets using Open API standards. Moreover, PIER nodes can create request models, which are views defined on the available datasets, possibly joining them.

PIER nodes rely on a powerful and agile oracle (namely, a component of the blockchain system that can read data stored externally), called Olinda, to create data services based on the OData protocol. Nonetheless, PIER nodes can recognize and import any dataset description that is Open-API compliant, as well as they can access any data service that supports the corresponding dataset RESTful API. By running configurable smart contracts, PIER nodes execute request models to retrieve data, and register all data requests (that is, the executions of the request models) in the distributed ledger, along with their responses.

In summary, in the PIER platform, blockchain ledgers are used to store: a decentralized catalog of datasets and request models; and an audit trail of all the data requests. Each participant is concerned only with the maintenance of its datasets and request models, which are automatically combined by the platform to compose the full catalog in the ledger. The PIER platform uses both public and private ledgers to enable flexible privacy control of the shared information. It explores the concept of dataspaces [1], such that the PIER platform provides information integration in a pay-as-you-go approach.

We developed the PIER platform using the JPMorgan Quorum software, and we are running a pilot in production since September of 2018 with the Brazilian financial regulators to support due diligence in authorization processes.

Currently, we are investigating further developments in the PIER platform, such as integrating off-chain credentials and datasets versioning in the decentralized catalog. Also, we are interested in exploring natural language processing and machine learning to classify and match datasets and their embedded data entities, as well as to automatically generate request models from datasets based on high-level user specifications.

References

1 Michael J. Franklin, Alon Halevy and David Maier. *From databases to dataspaces: a new abstraction for information management.* SIGMOD Record, vol. 34, n. 4, pp. 27-33, 2005.

4.22 Blockchain Empowered Drug Development Financing

Yong Tang (Univ. of Electronic Science & Technology – Chengdu, CN)

The drug developments are very risky business with very high failure rates and require massive investment. The procedure can last for near ten years before a successful drug is finally approved by FDA. The high risk makes the drug developers and investors less interested to invest in the early stages. To encourage investors to fund the underinvested stages requires innovative business model and platforms. In this talk, I'd like to introduce a blockchain empowered megafund for drug development financing. Using a blockchain-based special purpose vehicle (SPV), we get all stakeholders involved in drug development such as developers, SPV, regulators, institutional investors, retail investors, credit rating agencies, credit enhancers onto a platform. The data are shared, and the procedures are executed as smart contracts. All parties can enjoy better data sharing and enhanced services. More importantly, expensive management costs can be saved to allow better investment returns.

4.23 Turning a Vehicle Into an Economic Platform

Michael Huth (Imperial College London, GB)

Joint work of Kwok Cheung, Michael Huth, Laurence Kirk, Leif-Nissen Lundbæk, Rodolphe Marques, Jan Petsche
Main reference Kwok Cheung, Michael Huth, Laurence Kirk, Leif-Nissen Lundbæk, Rodolphe Marques, Jan Petsche: "Owner-Centric Sharing of Physical Resources, Data, and Data-Driven Insights in Digital Ecosystems", in Proc. of the 24th ACM Symposium on Access Control Models and Technologies, SACMAT 2019, Toronto, ON, Canada, June 03-06, 2019, pp. 73–81, 2019.
URL https://doi.org/10.1145/3322431.3326326

Consumer expectations and fierce market competition have led to margins becoming increasingly thinner for manufacturers of consumer and commercial vehicles. These actors realize now more than ever, that the value of their goods no longer rests on the basic functions they provide, but rather on the types and qualities of user experiences they can offer: extra horse-power on demand, ability to share usage, selling data streams to third parties – to name a few.

Increasingly, manufacturers are exploring ways to capture this value by turning a vehicle into a mini-economic platform that facilitates value exchange. Usage of that platform must be controlled so that value creation and consumption are neither impeded, nor corrupted, for the tenants that interact on it.

Our R & D in policy-based access control, distributed ledger technology, and embedded systems has led to the development of FROST Technology for fully programmable sharing ecosystems and flexible usage control on a vehicle's compute systems. FROST can thus provide consumers with novel, on-demand services whilst enabling manufacturers to tap into additional revenue streams.

4.24 Distributed Blockchain Systems across Distributed Data Centers

Dilip Krishnaswamy (Reliance Jio Infocomm Ltd., IN)

Emerging 5G and future networks will be realized leveraging programmable infrastructure that utilizes VMs and containers across hierarchical / distributed data centers. For transaction processing in such distributed deployments, distributed blockchain systems will need to be supported with consideration for data communication latency and bandwidth availability across these data centers. It would be interesting if a distributed blockchain system can be designed with lazy decoupling of blockchain ledgers that has a transaction throughput performance (tps) closer to a performance that is achievable in a local data center, while meeting the end-to-end latency constraint requirements across the distributed data centers over which the blockchain system is deployed. In particular, as edge data centers get deployed to provide support for latency sensitive applications (such as video streaming, Virtual and Augmented Reality applications, healthcare services, data privacy related services, financial applications, retail services, telecommunications services, etc.) at the edge of the network, transaction data will be produced at the edge where such transaction data will need closure in short time scales. Therefore a distributed producer-consumer blockchain framework with a scalable microservices-based approach is suggested in [1], where a producer blockchain sub-network commits transaction data locally, and then eventually commits the data to subscribing consumer remote ledgers with additional latency. Thus remote ledgers are only eventually consistent in such cases. For applications that are not latency-constrained, one can continue processing blockchain transaction data over a wider-area-network with reduced throughput. In general, based on the latency and throughput constraints that need to be met, one can choose to utilize such edge distributed ledger systems that synchronize lazily with remote blockchain ledgers, if desired.

References

1 D. Krishnaswamy at al, "A Microservices-based Virtualized Blockchain Framework for Emerging 5G Data Networks", IEEE Global Communications Conference (Globecom) 2019, Hawaii, December 2019.

4.25 Atomic Cross Chain Swaps

Eric Lo (The Chinese University of Hong Kong, HK)

Joint work of Eric Lo, Ziliang Lai, Lucian Ng, Sherman Chow, Yongyun Zhao

Since the birth of Bitcoins, thousand of new blockchains emerges. Allowing exchanges of digital currency and goods between blockchains helps users to enjoy benefits from different blockchains and improves the liquidity. To this end, we need a mechanism where multiple untrusted parties can exchange assets on different blockchains in an all-or-nothing manner, i.e., atomic cross-chain swaps. However, reaching consensus across different blockchains is challenging. Two outstanding issues are how to ensure all the blockchains 1) agree on swapping on not and 2) faithfully execute the swap protocol.

A native and common approach is running an exchange center to provide such service. However, such an approach violates the decentralized nature of blockchains since it places trust in the service provider. A common solution is to use smart contracts to escrow assets. Combining with hashed timelocks, a party holding a secret can decide swapping assets or not. However, hashed timelocks require synchronous clocks on different blockchains, which is missing in most blockchains. In this seminar, I introduced several solutions to attack this problem.
Collaborators: Lucien Ng, Sherman Chow, Yongjun Zhao, ZiLiang Lai

4.26 Blockchain Analytics

Murat Kantarcioglu (University of Texas – Dallas, US)

Joint work of Murat Kantarcioglu, Yulia Gel, Cuneyt Akcora
Main reference Cuneyt Gurcan Akcora, Yitao Li, Yulia R. Gel, Murat Kantarcioglu: "BitcoinHeist: Topological Data Analysis for Ransomware Detection on the Bitcoin Blockchain", CoRR, Vol. abs/1906.07852, 2019.
URL http://arxiv.org/abs/1906.07852

In this talk, we give an overview of the blockchain data analytics [3] where transactions recorded on blockchains such as Bitcoin can be represented as a heterogeneous graph [2] and then different graph patterns named chainlets [1] can be mined for predicting crypto-currency prices [1] to detecting ransomware activities [4]. In addition, we briefly discuss why some of the existing graph analytics techniques could not be directly applied for blockchain transaction graphs.

References

1 Cuneyt Gurcan Akcora, Asim Kumer Dey, Yulia R. Gel, and Murat Kantarcioglu. Forecasting bitcoin price with graph chainlets. In *Advances in Knowledge Discovery and Data Mining – 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018, Proceedings, Part III*, pages 765–776, 2018.

2 Cuneyt Gurcan Akcora, Yulia R. Gel, and Murat Kantarcioglu. Blockchain: A graph primer. *CoRR*, abs/1708.08749, 2017.

3 Cuneyt Gurcan Akcora, Murat Kantarcioglu, and Yulia R. Gel. Blockchain data analytics. In *IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018*, page 6, 2018.

4 Cuneyt Gurcan Akcora, Yitao Li, Yulia R. Gel, and Murat Kantarcioglu. Bitcoinheist: Topological data analysis for ransomware detection on the bitcoin blockchain. *CoRR*, abs/1906.07852, 2019.

4.27 Blockchain and Open Source Governance

Juho Lindman (University of Gothenburg, SE)

Early public blockchain cryptocurrency projects (such as Bitcoin and Ethereum) are licensed under open source licenses and governed openly in developer communities using governance mechanisms, practices and tools inherited from the open source world. Early governance ideals of these blockchain projects followed closely the governance of OSS such as Linux operating system or Apache Web Server. In my research I am investigating whether earlier open source software (OSS) research can help us to explain blockchain-related phenomena. OSS research offers several insights that may be reusable in Blockchain context regarding how to solve different kinds of tension between voluntary (developer) communities and profit-seeking commercial companies. The openness of artifact is obviously an interesting point of departure, but more critical questions may be related to guaranteeing the incentives of the different actors and matching divergent interests and misaligned incentives. The re-emergence of governance foundations and forks – splits of the development community – also raise questions where OSS may provide some analytical tools to raise the analytical clarity.

5 Findings

5.1 General Conclusions

Blockchain technology has many connections to database technology, yet the current situation is comparable to that in which the database field was in the 1980s: There were some commercial systems already, but users had to figure out for themselves how they could efficiently and effectively be used. As the seminar has shown, a number of important questions still have to be answered for blockchains, including but not limited to the following

- Blockchain scalability and performance w.r.t. transaction throughput is one of the core hurdles enroute to a wider application of blockchains. While commercial systems, e.g., those run by credit card companies, meanwhile achieve a throughput of 25,000+ transactions per second, blockchain systems can currently offer 15 to 20 transactions per second, due to their high verification effort. As a consequence, participants of a permissioned blockchain have to wait a long time for a transaction commit and hence for progress in the execution of a smart contract.
- Further development of the Proof of Work (PoW) consensus mechanism as well as other optimizations of finding consensus in the presence of transaction failures and w.r.t. scalability. Establishing consensus is of central importance since participants have to reach an agreement on the execution order of running transactions; this has to be identical for every party. The challenge here lies in the fact that while the (potentially large) participants in a Permissioned Blockchain are known, there is not necessarily trust among them, and some nodes may even be faulty (i.e., Byzantine errors have to be tolerated). However, blockchains do not always assume a complete lack of trust of actions among the participants, and so some do not get built on Byzantine-tolerating protocols.
- Languages for the specification of smart contracts are needed for an integration of blockchains into an existing enterprise IT. In addition, it is currently necessary for developers of smart contracts to use languages like Solidity which are low-level.

There is a lot of differences among the way words and terms are used, and what assumptions are made by the players in this field. Also, there is a wide variation in what aspects people are focused on: improving the performance of the protocols, integration with other data and computational platforms, understanding the security and fault-tolerance properties, applications, organizational aspects of managing the platforms, etc. It is also interesting to note that even when a trusted party does exist, there could be organizational constraints (budgets, mandates etc.) that lead to adopting a blockchain as a good architecture in practice.

5.2 More Blockchain Analytics

We saw that blockchains are used to manage both physical and digital assets, e.g., in finance, shipping, or energy (where I work a lot). Blockchains have both strengths and weaknesses, most notably performance, compared to other technologies for managing distributed data and transactions, e.g., database systems. It is thus interesting to investigate the optimal technology mix for certain types of analytics applications. Specifically, it is very interesting to be able to analyze the large amounts of data on blockchains. Research questions include the following:

- How can blockchains be optimally combined with (existing) database and analytics technology for different types of analytical workloads?

- How can data in blockchains be analyzed in powerful and scalable ways, like for data in normal databases?
- Which new, specific types of analyses are needed for blockchain data?

5.3 Virtual Assets

Blockchain technology can be used to establish a whole class of virtual assets, such as individual data, by providing both the "trust" and "incentive" necessary to foster a self-growing value ecosystem. In particular, one can explore more on using blockchain to solve the bottleneck in today's data-driven economy – how we initiate and push along "Symphony Protocol" to create an ecosystem that unlocks personal data for democratized and personalized intelligence, with privacy by design. The domain presents a wealth of interesting research questions, such as data pricing and trading.

5.4 Areas for Future Work

Among the activities for future work in the area of (permissioned) blockchains, participants of the seminar suggested the following:

- Foundation of an Academic Research Special Interest Group: The goal of this group is to be a forum for academic researchers in Hyperledger. We want researchers that are interested in Hyperledger or Hyperledger-related topics to be able to interact and collaborate on problems. This might take the form of presentations, discussions, or collaborative working sessions. We also will incorporate bidirectional communication with developers and engineers.
- Writing of a "Blockchain Manifesto" which helps clarifying the terminology used in this area and sorts out as well as organizes the main directions of development by which the field is characterized. Several participants have expressed their interest in contributing to such an endeavor.

Acknowledgements We want to thank the Dagstuhl staff for providing an environment that truly encourages scientific exchange and discussions and that cares for the guests in an unmatched way.

Participants

- Divyakant Agrawal
University of California – Santa Barbara, US
- Alysson Neves Bessani
University of Lisbon, PT
- Jeeta Ann Chacko
TU München, DE
- Lei Chen
HKUST – Kowloon, HK
- Mariano P. Consens
University of Toronto, CA
- Tien Tuan Anh Dinh
Singapore University of Technology and Design, SG
- Alan Fekete
The University of Sydney, AU
- Michael J. Franklin
University of Chicago, US
- Vincent Gramoli
The University of Sydney, AU
- Krishna P. Gummadi
MPI-SWS – Saarbrücken, DE
- Michael Huth
Imperial College London, GB
- Hans-Arno Jacobsen
TUM, DE & Univ. Toronto, CA
- Murat Kantarcioglu
University of Texas – Dallas, US
- Srinivasan Keshav
University of Waterloo, CA
- Shahan Khatchadourian
Toronto, US
- Dilip Krishnaswamy
Reliance Jio Infocomm Ltd., IN
- Juho Lindman
University of Gothenburg, SE
- Eric Lo
The Chinese University of Hong Kong, HK
- Alexander Löser
Beuth Hochschule für Technik – Berlin, DE
- Dumitrel Loghin
National University of Singapore, SG
- Bernhard Mitschang
Universität Stuttgart, DE
- C. Mohan
IBM Almaden Center – San Jose, US
- Hart Montgomery
Fujitsu Labs of America Inc. – Sunnyvale, US
- Pezhman Nasirifard
TU München, DE
- Beng Chin Ooi
National University of Singapore, SG
- Torben Bach Pedersen
Aalborg University, DK
- Dennis Przytarski
Universität Stuttgart, DE
- PingCheng Ruan
National University of Singapore, SG
- Gabriela Ruberg
Banco Central do Brasil – Rio de Janeiro, BR
- Mohammad Sadoghi Hamedani
University of California – Davis, US
- Yong Tang
Univ. of Electronic Science & Technology – Chengdu, CN
- Gottfried Vossen
Universität Münster, DE
- Li Xiong
Emory University – Atlanta, US
- Feida Zhu
SMU – Singapore, SG

Report from Dagstuhl Seminar 19262

Astrographics: Interactive Data-Driven Journeys through Space

Edited by

Alyssa A. Goodman[1], Charles D. Hansen[2], Daniel Weiskopf[3], and Anders Ynnerman[4]

1 **Harvard-Smithsonian Center for Astrophysics, US,** agoodman@cfa.harvard.edu
2 **University of Utah – Salt Lake City, US,** hansen@cs.utah.edu
3 **Universität Stuttgart, DE,** daniel.weiskopf@visus.uni-stuttgart.de
4 **Linköping University, SE,** anders.ynnerman@liu.se

Abstract

This report documents the program and outcomes of the Dagstuhl Seminar 19262 "Astrographics: Interactive Data-Driven Journeys through Space". The seminar consisted of introductory talks, which are presented first in this documents, followed by discussions in break-out groups whose results were reported back to all participants after each break-out session.

Seminar June 23–26, 2019 – http://www.dagstuhl.de/19262
2012 ACM Subject Classification Applied computing → Astronomy, Human-centered computing → Visualization, Human-centered computing → Visualization toolkits
Keywords and phrases astronomy, data visualization, immersive environments, planetariums, space exploration
Digital Object Identifier 10.4230/DagRep.9.6.95
Edited in cooperation with Alexander Bock

1 Executive Summary

Alexander Bock (Linköping University, SE & University of Utah, US)
Alyssa A. Goodman (Harvard-Smithsonian Center for Astrophysics, US)
Charles D. Hansen (University of Utah – Salt Lake City, US)
Daniel Weiskopf (Universität Stuttgart, DE)
Anders Ynnerman (Linköping University, SE)

For the majority of human existence, the visual language has been successfully used to communicate complex ideas that span across borders of knowledge, experience, age, gender, culture, and time. These aspects also make it an effective form of expressing workflows in scientific data analysis as well as the communication of scientific discoveries to broad audiences. The Dagstuhl Seminar 19262 brought together researchers from computer science, content producers, learning and communication experts, and domain experts from astronomy and astrophysics to define the emerging field of interactive visualization of space exploration and astronomy, referred to as Astrographics. This seminar played an important role in the ongoing process of removing the clear division between using visualization to enable scientific discoveries by subject-matter experts (exploratory visualization) and using visual representations to explain and communicate the results of such exploratory science to a greater, general audience (explanatory visualization). Designing the available visualization tools to serve both roles at the same time increases the overlap between these two aspects of visualization and allows scientists to better explain their findings and, at the same time, enables the general public to use similar tools for their own, guided, discovery and actively participate in the scientific process. The field of astronomy and astrophysics has been at the forefront of this process since the beginning as it is a primary example of a domain in which exploratory and explanatory visualizations have served important but distinct roles. For this reason, astrographics was chosen as the domain in which to explore the challenges and opportunities that arise when combining exploratory and explanatory techniques. The bulk of work in this seminar occurred in focussed break-out sessions that reported their findings back to the group and opened up the topics for joint discussions. Topics of these break-out sessions included discussions on better integration of software tools, improvements of analysis tools, preparing astrographics software packages to improve the quality of public presentations, the ability of sharing presentations both in spatially distant locations as well as saving them for later playback. Finally, there was a working group to work on a decadal white paper for astronomy [1].

References

1 Jacqueline K. Faherty, Mark SubbaRao, Ryan Wyatt, Anders Ynnerman, Neil deGrasse Tyson, Aaron Geller, Maria Weber, Philip Rosenfield, Wolfgang Steffen, Gabriel Stoeckle, Daniel Weiskopf, Marcus Magnor, Peter K. G. Williams, Brian Abbott, Lucia Marchetti, Thomas Jarrrett, Jonathan Fay, Joshua Peek, Or Graur, Patrick Durrell, Derek Homeier, Heather Preston, Thomas Müller, Johanna M Vos, David Brown, Paige Giorla Godfrey, Emily Rice, Daniella Bardalez Gagliuffi, Alexander Bock. IDEAS: Immersive Dome Experiences for Accelerating Science. arXiv preprint arXiv:1907.05383, 2019

2 Table of Contents

Break-out Groups

3 Overview of Talks

3.1 Developing Tools for Increased Access to Astrophysics in OpenSpace

Brian Abbott (AMNH – New York, US)

I attended the AstroGraphics meeting in order to discuss the various ways in which we can advance OpenSpace for all types of users. Developing and disseminating astrophysics data within OpenSpace is the overarching goal, for researchers and the public alike, and I wish to discuss the methods to achieve these goals. The result of these advancements will open up access to scientists, who will explore their data in new, insightful ways, and to the public, who will be inspired by their immersion in the universe and enhance their interest in STEM fields.

Presentation Tools

Methods of creating so called Show Kits that include information for training purposes as well as for use in production and storytelling is a large part of opening access to power users. Kits that include flightpaths, documentation, annotationed flights embedded within interactive timelines and overlaid in a video version of the content, and an audio component will serve to provide tools for production and live programming, but will also be critical for training presenters.

Showlettes

Show Kits can be composed of individual showlettes, which are bite-sized stories that contain a beginning, middle, and end, but are confined to one action of concept. For example, exploring the Apollo 11 landing site may be one showlette. Or, the radio sphere with the exoplanets may be one showlette. These showlettes could be combined to form longer shows that take on a broader narrative. Showlettes could be tied together by their beginning and end points, or perhaps these points are adjustable. Some OpenSpace development would be required to make this possible.

Authoring Tools

Authoring Tools are an immediate need for OpenSpace in order to increase access for power users and data holders. The goal is to transform OpenSpace from "under-the-hood," script-based access to more of a "plug-and-play" model, where the user does not need to know or interact with the underlying scripting language. A data questionnaire could be developed to accommodate the import process and generate the underlying scripts on the fly. Tools such as these will cater to all levels of groups who wish to go beyond the distributed offerings of OpenSpace and build their own datasets, and will lower the bar for researchers and power users to use openSpace to interact with their own data and develop stories around those data.

Domecasting Network

In order to build a more cohesive community, we want to develop a streaming network so that users are aware of streaming and domecasting events and can plan for them. Building a domecasting network will allow high-end scientific data to reach audiences all over the world. And, such a network will foster more communication between users and enhance the resources available for everyone.

3.2 Astrographics: Introduction of NAOJ and OAO / IAU

Hidehiko Agata (NAOJ – Tokyo, JP)

Introduction of 4D2U project: Our universe encompasses all of space and time. The most immense cosmic structures – galaxies gathered into clusters and superclusters – dominate the universe everywhere we look. Astronomers constantly work to understand these structures bit by bit. The Four- Dimensional Digital Universe (4D2U) project is an outgrowth of their search for cosmic understanding. The project's goal is to visualize astronomical data in a way that helps watchers feel as if they are witnessing the unfolding of our universe. "Four dimensions" refer to the three dimensions of space and the one dimension of time embedded in their data. "Digital" refers to computer graphics visualizations of digital data. The resulting acronym is "4D2U", and it is also astronomy's way of saying "4D to you". The 4D2U is a long-term project by the National Astronomical Observatory of Japan (NAOJ). We plan to continue improvements in our visualization data library and further development of new visualization platforms, including domed and portable theaters. We hope you'll have the opportunity to experience 4D2U. The project hosts public showings every two months at the NAOJ headquarters in Mitaka, Japan.
https://4d2u.nao.ac.jp/english/index.html
https://4d2u.nao.ac.jp/html/program/mitaka/index_E.html

Introduction of OAO/IAU: The IAU Office for Astronomy Outreach (OAO) is IAU's hub for coordinating its public outreach activities around the world. The OAO coordinates and and supports worldwide efforts to enhance public knowledge, appreciation and education of astronomy and related sciences. The OAO promotes public awareness of the IAU activities and coordinate and manage the IAU international outreach campaigns. The OAO maintain and coordinate the IAU network of National Outreach Coordinators (NOC), who will be responsible for the implementation of the proposed IAU outreach initiatives at national level and for maintaining the relationship with the national communities of amateur astronomers. The OAO act as the central information hub for disseminating IAU related public outreach activities around the world, as a facilitator for best practices, providing guidance and delivering regular information about ongoing activities in astronomy communication.
https://www.iau.org/public/
https://www.iau-100.org/

3.3 Introductory Presentation – Linköping University

Emil Axelsson (Linköping University, SE)

I am a Research Engineer at Linköping University and one of the software developers of the OpenSpace software. Me and my colleagues have long had the vision of building a tool that can be equally helpful for public science communication as it is for scientific visualization and visual data analysis. While some of these visions have been turned into reality over the years, the aspect of using our virtual reality representation of the universe as a science tool have been one of the more fuzzy ones. After this Dagstuhl seminar, I no longer feel that this goal is fuzzy: Instead, it is clear that there are several of us who share this vision, and that our ideas are coming together into a new way of using astronomy tools and visualization software together to analyze data and communicate science.

When I arrived to this Dagstuhl seminar, I was excited about the opportunity to discuss connecting OpenSpace to analysis tools used by astronomers. It was amazing to meet the brains behind prominent astronomy software packages, such as Glue, Astropy, and Aladin. I was thrilled by how we shared this common vision and were able to sit together for a few hours and create a first proof of concept of how we link our tools together. I'm very excited to continue working together with the people I've met here, to create new ways to visualize astronomy data for science and communication!

3.4 Bringing together professional astronomy and EPO visualization software tools

Thomas Boch (Université de Strasbourg – Strasbourg, FR)

I am a research engineer at CDS (Centre de Donnés astronomiques de Strasbourg) since 2002. My activities for 17 years have involved the development of software tools and systems to enable efficient access to the reference data we host at CDS, through Virtual Observatory . In the last decade and with the advent of large astronomical surveys like Gaia or Pan-STARRS, the Big Data has become another dimension to take into account. We have developed a multi-resolution approach to this challenge, supported by the HiPS (Hierarchical Progressive Surveys) format, which has gained traction as more than 600 astronomical surveys (mostly images) are available as HiPS collections (as of June 2019).

While developing the Aladin applications, I was also attracted by the nice aesthetics presented by the planetarium softwares. This seminar has been a great opportunity for me to learn about up to date tools and techniques used for education and public outreach in domes and planetariums. I have been particularly impressed and inspired by the potential of OpenUniverse, especially for the exploration of planetary surfaces.

During the break-out sessions, I created a prototype connecting Aladin Lite to OpenSpace, as a proof of concept. This has showed some interesting potential and interest. More generally, collaboration between the professional astronomy visualisation and the EPO/planetarium communities is of mutual benefit and I feel very grateful for having the opportunity to participate in this seminar and make useful connections with people I hope to work with in the future.

3.5 OpenSpace – Visualizing the universe

Alexander Bock (Linköping University, SE & University of Utah, US)

I presented an overview of the OpenSpace platform that aims to provide an astrovisualization toolkit to visualize the entire universe. These visualizations cover a large range of datasets as well as scales, ranging from millimeter sized rocks on Mars all the way out to the edge of the observable universe. The software is catered to serve three distinct target audiences: The public can be reached both through the use of presentations as well as self-directed discovery in exhibition spaces. A second target audience is visualization researchers that can use a common software framework to conduct novel research. A third target audience are domain scientists that can use the same software to conduct their own research as is used for public presentations, for example planetarium domes. I presented a selection of previous works that have already been published or shown to the public. Many of the presented examples were the result of Master thesis works, of which I exemplified three in the initial presentation. These examples covered the flyby of the New Horizons spacecraft at Pluto in 2015, volumetric rendering of Coronal Mass Ejections from the Sun, all the way to representations of the traffic on the NASA Deep Space Network. Concluding the talk, I showed institutions that already make use of OpenSpace in public programming and provide food for through on how we can continue the development of astrographics tools to increase their reach, particularly in the case of interaction techniques.

3.6 David Brown

Dave Brown (Microsoft Research – Redmond, US)

My research interests focus around data visualization in plan and mixed reality, using devices large and small, and with an emphasis on storytelling. I have also long had an interest in astrographics, having met Curtis Wong through a project I was working on to visualize the solar system and its surroundings on table-top form factor devices. I subsequently worked for Curtis at Microsoft Research, and got to know other members of the team behind WorldWideTelecsope, and key visionaries in this area such as Alyssa Goodman. This Dagstuhl was a perfect intersection between my current research in data-driven storytelling and desire to do more with astronomical data. It gave me both an opportunity to share some of my work with the other participants, and for me to learn from experts in the astrographics and full-dome experiences domain. After so many interesting discussions and much brainstorming with the other attendees, I am convinced that full-dome data-exploration and storytelling is an un-tapped opportunity for any data-driven research and communication, and that it overlaps with my current research in exploring how collaboration and mixed reality can benefit data visualization. This Dagstuhl was therefore an unprecedented opportunity for me to forge new networks with people with similar interests but who worked in fields that I would not otherwise have had the opportunity to meet. I am looking forward to future work with these new-found colleagues.

3.7 Exploring the Galaxy Using Virtual Reality

Melvyn Davies (Lund Observatory, SE)

I am a dynamic theoretical astrophysicist who studies the evolution of exotic objects (planetary systems, black holes and explosive events) in crowded places. How does one explore the rich diversity of stellar systems? This is an enormous visualisation challenge. In particular in Lund we are exploring how one can use Virtual Reality to explore rich stellar data sets. Using Unity we are learning how to interact with point cloud data: the positions and velocities of an ensemble of stars where each point also possess other attributes such as stellar age and chemical abundance. Our goal is to employ this technology for scientific discovery with observational data from the ESA Gaia astrometric mission together with outputs of computer simulations of galaxy formation. During the seminar, we discussed a broad range of important issues concerning the use of Virtual Reality and Augmented Reality. The diversity of the knowledge present was enormously helpful in understanding the future prospects and a way forward.

3.8 Planetariums as a Research Tool

Carter Emmart (AMNH – New York, US)

For twenty years I have been with the American Museum of Natural History (AMNH) helping to craft public demonstration of our state of knowledge about the universe using data visualization. The planetarium as a hemispherical projected image of the night sky with its attendant motions, movements of celestial objects, simulated views from different latitudes and times across the precession cycle has from the start been accurate science visualization, portrayed two dimensionally. Research trends in immersive data visualization explored in the last two decades of the twentieth century became possible to display in planetarium domes at the turn of the Millennium when AMNH renovated the Hayden Planetarium, recasting its presentations from the tradition of Sky Shows toward new programming termed Space Shows using data visualization to portray the universe accurately in three dimensions. In both cases of the traditional 2D and modern 3D display capabilities of the planetarium, the primary purpose has been to service public understanding through informal education. Planetariums have been on the receiving end of science as an effective demonstration tool. While modern capability of 3D interactive display affords data exploration, aspirations of using planetariums as scientific tools of discovery have to date remained a noble vision. The June, 2019 Dagstuhl Astrographics seminar afforded a very unique, narrow time window for a potent mix of talent across the field that forged an exploration of data flow using open tools (OpenSpace, Glue, CDS Aladin) bridging scientific investigation directly into planetarium display with its capacity for audiences numbering in the hundreds. I raised a question at the conclusion of Astrographics if what we were seeing was useful to the scientists represented in the meeting, and their resounding response was yes is was. Going forward, this new beginning will develop this trend further which holds great potential for the field of data visualization supporting scientific investigation in planetariums worldwide.

3.9 Introductory Presentation – American Museum of Natural History

Jackie Faherty (AMNH – New York, US)

I am a senior scientist and senior education manager jointly in the departments of Astrophysics and Education at the American Museum of Natural History. Scientifically I am interested in exploring questions of the tail end of the star formation process. I am specifically focused on studying the fundamental properties of low mass stars, brown dwarfs, and exoplanets. My research is greatly influenced by big data surveys such as ESA's Gaia observatory. In addition I have found that large astronomical data sets are best investigated scientifically using visualization tools as an aid. This has brought me to the Astrographics workshop as I am in constant search of software packages that will help with interpreting scientific content. As I work in the planetarium world, I am exposed constantly to visualization tools for the purpose of education – reaching the general public. I see an important collision in today's technologically advanced world between the scientific agenda of researchers in the era of big data and the sophisticated tools being used today to educate and inspire the general public.

3.10 CosmoScout VR: Interactivity and Immersion for Space Data Exploration and Mission Planning

Andreas Gerndt (DLR – Braunschweig, DE)

Although the German Aerospace Center (DLR) is not directly addressing astrophysics, many institutes of DLR are on earth system modeling, planetary research, atmospheric research, space mission planning, space operation, on-orbit-servicing, space situational awareness, air-borne observation, and sensor-data related applications. The observation and simulation data can become extremely large. Additionally, heterogeneous data exploitation requires software solutions which are distributed, scalable, and reliable. On the Dagstuhl seminar, I could give in insight into DLR's manifold space activities and the interactive data exploitation software we have been developing. Already 10 years ago, the DLR Institute of Simulation and Software Technology realized the need to process the huge High-Resolution Stereo Camera (HRSC) dataset on a global scale on a sphere in virtual environments. HRSC is a DLR camera sensor on the Mars Express orbiter capturing stripes of the surface of Mars with extraordinarily precise specification of slopes. An accurate data exploration requires a distortion free mapping also on the poles and appropriate tools for measuring, annotation, and navigation. To close gaps between stripes, arbitrary sensor datasets can be merged on the fly to one data product. Virtual reality devices have been integrated to sense the height-field for accurate measurements and to deform the terrain interactively for reconstructing 3D fault displacements. Autonomous crater detection based moon lander required simulators with physical-based Moon rendering. And for Earth observation data, the interactive terrain renderer has been extended with approaches for atmospheric volumetric and time-depended data visualization. Climate research datasets exceeds Peta-bytes of data, so that the framework has been extended to remote high-performance post-processing on supercomputers. And finally, space mission planning demanded the visualization of the whole solar system with accurate time-dependent trajectory specifications based on SPICE kernels. High-resolution star maps allow space system simulators which rely on star trackers. On-orbit servicing simulation of spacecraft requires photo-realistic rendering of the scenario in space. All these features are now available in the open-source software framework CosmoScout VR. It works as a stand-alone application. With its plug-in concept, it additionally enables the extension of arbitrary functionalities. But as it comes as a set of libraries, tailored applications for specific purpose can be implemented as well. With the basic VR software layer, it always scales from single laptop application to CAVE-like immersive multi-projection systems. And finally, it is platform independent and supports Windows and Linux operating systems. Thus, CosmoScout VR is available for all who are interested in a holistic solution to explore the solar system in 3D and time. Its approaches can clearly contribute to the astrophysics community. It was great to meet all the people at Dagstuhl to discuss their requirements in visualizing astrophysical data and to figure out opportunities of interactive exploration.

3.11 Astrographics, Exploranation, and Dagstuhl

Alyssa A. Goodman (Harvard-Smithsonian Center for Astrophysics, US)

For the first many years of my professional career, I was lonely living at the intersection of astrophysics and data visualization. Even when I attended my first "astro-visualization" meeting back at KICP in Chicago in 2005, I felt like an outsider, as I was not part of the planetarium community where most research scientists (like me) felt "visualization" and "visual communication" belonged. At that Chicago meeting, I presented my group's first attempts to use Medical Imaging software on Astronomy data, under the "Astronomical Medicine" project at Harvard's Initiative in Innovative Computing. At the same meeting, Curtis Wong presented his ideas about how to create what ultimately became WorldWide Telescope (WWT, worldwidetelescope.org) , which I was later honored to help Curtis and Jonathan Fay create.

By now, over a decade past the release of WWT, my collaborators and I have created the open-source glue visualization environment (glueviz.org), which lets users "glue" together data sets, visualizations, and software tools, including WWT. Interest in, and usage of, glue from all quarters of astronomy and the broader scientific world is on the rise, which thrills me. My colleagues in Astronomy have, over just the past five years or so, rapidly come to hunger for visualization tools and innovations, thanks to the advent of ever-larger and more complex data sets, as well as more sophisticated statistical approaches, all of which essentially require visualization to understand.

Here at Dagstuhl, I witnessed a transformation. Before the meeting, WWT was the only full-featured "sky browser" available in glue via its plug-in architecture. But now, after just 2 days of discussion and hacking, the research and planetarium communities in astronomy have come together to link up their "exploratory" and "explanatory" tools into amazing "exploranation" mash-ups that will transform how the public experiences real data, and how and where researchers view their data. For example, as a result of our extensive discussions and real-time hacking/collaboration, Aladin (a European program widely used in Astronomy research), as well as glue, can now both communicate with OpenSpace (an immersive 3D data-driven visualization environment used in the planetarium community). My astronomy colleagues and I have already–just in the last day of the Dagstuhl meeting–begun to use these newly-linked tools to explore and discover new, 4D-spacetime, relationships between stars and gas in the Milky Way.

Back-and-forth travel from "exploration" to "explanation" in visualization holds tremendous promise for both professionals and the wider public's understanding of science, and this first Astrographics Dagstuhl Seminar will stand out in the future as key to enabling this two-way travel!

3.12 Plenary Astrographics Presentation

Charles D. Hansen (University of Utah – Salt Lake City, US)

I am a faculty member of the Scientific Computing and Imaging Institute at the University of Utah. One of the primary application foci of the SCI Institute has continued to be biomedicine; however, SCI Institute researchers also address challenging computational problems in a variety of application domains, including materials, manufacturing, defense, and energy. SCI Institute research interests generally fall into the areas of scientific visualization, scientific computing and numerics, image processing and analysis, and scientific software environments. SCI Institute researchers also apply many of the above computational techniques within their own particular scientific and engineering subspecialties, such as fluid dynamics, biomechanics, electrophysiology, bioelectric fields, parallel computing, inverse problems, and neuroimaging. A particular hallmark of SCI Institute research is the development of innovative and robust software packages, including the SCIRun scientific problem-solving environment, Seg3D, ImageVis3D, VisTrails, ViSUS, Cleaver, and map3d. All these packages are broadly available to the scientific community under open-source licensing and are supported by web pages, documentation, and user groups.

My research area is in scientific visualization. Volume visualization has been one of the areas where I have made contributions. Bring better illumination models to bring out more structure in 3D data, providing multi-dimensional transfer functions to better examine data and methods for large-scale rendering and analysis. I have incorporated research results into large-scale software projects. One such project is FluoRender. FluoRender is an interactive rendering tool for confocal microscopy data visualization. It combines the rendering of multi-channel volume data and polygon mesh data, where the properties of each dataset can be adjusted independently and quickly. The tool is designed especially for neurobiologists, allowing them to better visualize confocal data from fluorescently-stained brains, but it is also useful for other biological samples.

In Astrographics, I have made contributions from methods to visualize galaxy formation in collaboration with Elena D'Onghia (UWisc/Harvard) to methods for visualization of magnetic pole reversal with Ben Brown (UWisc). For the future of Astrographics, the combination of techniques, methods, and tools to aid in the exploration and explanation of scientific data is important. I participate in the OpenSpace project which brings NASA mission, astronomy data, and astrophysics data to the public through planetarium presentations.

3.13 Tom Kwasnitschka

Tom Kwasnitschka (GEOMAR Helmholtz Centre for Ocean Research Kiel, DE)

As a marine geologist, my research focuses on the exploration of oceanic volcanism down to extreme depths of several kilometres. Since most of such research is carried out using remotely operated diving robots, we heavily rely on remote sensing, manipulation and powerful visualization tools to build situational awareness of abyssal environments.

My work aims to reflect the entire value-added chain of the scientific process. Formulating research questions based in physical volcanology, I entertain projects for the design and construction of deep ocean cameras and optical sensing equipment such as LED lighting. My processing method of choice is photogrammetry because the high-resolution, color re-creation of underwater outcrops caters to the first-person based interpretation schemes of classic field geology. In order to do so, the photogrammetric models need to be adequately viewed and manipulated in spatially immersive environments, which my group designs, builds and operates. We focus on domes, drawing from the developments in the planetarium field throughout the last 15 years, as domes have a number of advantages when it comes to mult-viewer situations.

With a background in the planetarium business, my ties to astrovisualization are explained. I believe that even localized visualizations should be nested in the global perspective of a virtual globe, and that such a virtual globe is naturally part of a system that also simulates the proximal and distal cosmic phenomena. Therefore, I try to lobby for a broadening of scope and capabilities of astrovisualization software packages such as the ones used for domes and planetariums, especially towards the accomodation of earth science based data sets. This must also come with a refinement of tools to allow manipulation and agile interaction with the data behind a simulation, since exploratory visualization should always offer a quantitative, added value to the scientific process. Thus, and lastly, it is important to install measures to quantify the (hopefully positive) impact of visualization efforts to sustain and broaden its acceptance.

3.14 OpenSpace Amplification for VR and Other Disciplines

David H. Laidlaw (Brown University – Providence, US)

My research interests are in supporting scientific research with visualization and computation. I study efficacy of visualization methods and devices, with a recent focus on virtual reality displays. In this meeting I would like to explore how to make astrographics tools like OpenSpace work on more visualization devices. I am also interested in how ocean and bay science on earth might be supported by astrographics tools and how new data types (eg, lidar and ocean/bay sampling) might be incorporated.

3.15 Marcus Magnor

Marcus A. Magnor (TU Braunschweig, DE)

Humans have been fascinated by astrophysical phenomena since prehistoric times. But while the measurement and image acquisition devices have evolved enormously by now, many restrictions still apply when trying to estimate the three-dimensional structure of extraterrestrial objects. The most notable limitation is our confined vantage point, disallowing us to observe distant objects from different points of view. In an interdisciplinary German-Mexican research project with Wolfgang Steffen, partially funded by German DFG and Mexican CONACyT, we evaluated different approaches for the reconstruction of plausible three-dimensional models of planetary nebulae. The team was comprised of astrophysicists working on planetary nebula morphology and computer scientists working in the field of constrained reconstruction and physically correct rendering of astrophysical objects.

3.16 Thomas Müller

Thomas Müller (Haus der Astronomie – Heidelberg, DE)

The main task of the Haus der Astronomie is public outreach for the Max Planck Institute for Astronomy and public outreach of astronomy in general. We offer regular series of lectures, guided tours, workshops, teacher training and a lot more. My research interests cover relativistic visualization and any other kind of visualization in astronomy and astrophysics. My daily work is to help MPIA scientists in any questions of visualization, develop standalone exploratory and explanatory visualization applications for scientists, public outreach, and our in-house planetarium.

3.17 Josh Peek

Joshua Eli Goldston Peek (Space Telescope Science Inst. – Baltimore, US)

As an data-oriented astronomer working on diffuse structures in the universe, my world is ruled by a number of principles. Firstly, collaboration. Since the data I use is not dominated by a single data collection mode, but rather comes from radio, optical, simulation, space, ground, IFU, spectral, etc sources, I often work closely with experts to achieve data synergy. Each data set has its own quirks limits and biases. Secondly, dimensionality. As derived from the first, each data set provides incredibly high dimensional data, and combined the dimensionality is incredible. Lastly, the visual. The structures we explore are not always entirely encapsulated by points, but rather are extended and cannot be simply shown in scatter plots.

All of these things drive me toward astrographic analyses; using advanced visualization tools to explore complex overlapping data sets and exploring them with colleagues.

3.18 Lucian Pleasea

Lucian Plesea (ESRI – Redlands, US)

I have a long history of working with remote sensing imagery and data, initially for NASA and later for Esri, with the main goal of making such data available to the public at large. Visualization such data properly is very important, the volume and scale of the available datasets only make this visualization more challenging. It is only natural that planetariums are some of the best venues for astro-graphics, and I have been collaborating with NASA and AMNH to achieve a seamless and rapid infusion of planetary imagery data into planetarium visualization. The OpenSpace software is a remarkable advance in the astro-graphics field, the features and capabilities already exceed many commercial alternatives. Since it is open source and has relatively modest minimum hardware requirements, it is widely available to anyone interested. This positive impression of OpenSpace was widely shared by the workshop participants. Since I am also a member of the OpenSpace project advisory board, it was great to see how well received the OpenSpace development effort is. The workshop at Dagstuhl represented a great opportunity for me to meet with specialists working directly in the astro-graphics field, and to better understand the current state of the art, with its challenges and limitations. As a result of the workshop, I have a list of new action items I set for myself, and new motivation to continue to be an active contributor to astro-graphics.

3.19 Sebastian Ratzenböck

Sebastian Ratzenböck (Universität Wien, AT)

I am a PhD student at the data science platform at the University of Vienna focussing on extracting physical knowledge from the Gaia data set by employing machine learning methods. In a quest to better understand the structure formation of the milky way I am currently tailoring clustering algorithms to more robustly extract open clusters. For me data visualisation and data analysis, such as the clustering process, share a common basis as both tools are trying to make sense of the great unknown hidden in the data which essentially consists of numbers stored somewhere on hard drives. This notion of discovery and making sense of the data is now becoming more and more relevant as the astronomy community is facing massive amounts of data from survey like Gaia or Pan-STARRS1. Moreover, especially in astronomy where the search for structure involves the position space the focus on visualisation is a crucial part of the analysis and eventually telling the story behind the findings.

This seminar has been a first chance for me to get immersed into the joint field of astronomy and visualization and to meet the people on the forefront of that community. People here have in just two days managed to join softwares which further progresses and facilitates the ability to do exploratory data analysis. Furthermore, we discussed the fundamental structure of the data analysis pipeline from data discovery, data accessibility, transparent analysis especially regarding VIS tools and eventually novel analysis techniques such as Machine Learning in the context of Astrographics.

3.20 Thomas Robitaille

Thomas P. Robitaille (Aperio Software – Leeds, GB)

I am a scientific software developer working on a variety of open source projects related to visualization and astrographics. One of my key interests is making it possible for researchers to have access to tools for interactive data analysis, and in particular using different tools together seamlessly. At this workshop, I learned about a variety of advanced visualization tools being developed, and decided to work on developing a proof-of-concept of how to use the OpenSpace application with the glue data visualization software. This will allow researchers to carry out advanced data analysis, linking datasets together, and exploring interesting subsets of data while also being able to visualize these data and subsets in realistic rendering of the universe.

3.21 Introductory Presentations

Filip Sadlo (Universität Heidelberg, DE)

My research interests in astrophysical visualization range from the early Universe to planet formation. I would like to employ topological techniques to analyze the structure of the early Universe and planet formation, with a particular focus on its multi-scale nature. Besides the growing range of scales, I see challenges for future astrophysical visualization research in merging and completing heterogeneous data and a tighter coupling of the visualization concepts to the governing physical laws.

3.22 Wolfgang Steffen

Wolfgang Steffen (Universidad Nacional Autonoma de Mexico – México, MX)

Volume rendering of astronomical nebulae in interactive dome software is still underdeveloped with all the rest. We have built software (Shape) that allows to generate volumetric 3-D models, but integrated them into presentation software has been a problem in various ways: rendering quality, size limits and opacity rendering whenever there are non-volumetric objects present, too. Solutions to these problems and finding data format standards would be extremely helpful once hardware will become common place that is able to support high resolution volumetric models.

3.23 Introductory Presentation – Planetarium of the Natural History Museum Vienna

Gabriel Stöckle (Naturhistorisches Museum Wien, AT)

I am responsible for the digital planetarium in the Natural History Museum Vienna. The museum is one of the largest and most important natural history museums in the world with a unique collection. The planetariums purpose is to be an addition to the museum and to visualize and explain nature. We are not really doing scientific visualization yet but work rather as a planetarium with live presentations and fulldome cinema. Now with modern visualization techniques, I think a planetarium can become a tool to visualize scientific data interactively. We have Open Space installed and running (with support from the Gene Payne from the Scientific Computing and Imaging Institute in Utah) and hope to implement modern visualisation technologies in the dome.

My personal background is that I was working for the Astronomisches Rechen-Institut in Heidelberg, Germany promoting Virtual Observatory Tools and Grid Computing before I became responsible for the planetarium. Now with Open Space I found an open source software to visualize astronomical data on the dome, tested and installed it our dome and therefore was invited to come to Dagstuhl. My hope was to improve it into an tool to present interactively data on the dome.

Here in Dagstuhl, I met amazing people and we made great advances in visualizing astronomical data with the help of the developers of glue (Thomas Robitaille) and aladin (Thomas Boch). They were able to connect the Open Space Software with glue. They were able to select subsamples of data interactively and visualize these subsamples interactively in open space, which means also on the dome. I hope that this interactive data selection and presentation technique can also become a tool for researchers in the dome.

Next to this I could learn a lot about presentation techniques listening to the presentation of Carter Emmart from the American Natural History Museum in New York. I learned a lot and will use his presentation of the moon and mars as an inspiration for our own liveshows.

References

1 F. Bonnarel, P. Fernique, O. Bienaymé, D. Egret, F. Genova, M. Louys, F. Ochsenbein, M. Wenger and J. G. Bartlett. The ALADIN interactive sky atlas. Astron. Astrophys. Suppl. Ser, 143(1):33–40, 2000
2 T. Boch, P. Fernique. Aladin Lite: Embed your Sky in the Browser. Astronomical Data Analysis Software and Systems . 485:277, 2014

3.24 Data to Dome

Mark Subbarao (Adler Planetarium – Chicago, US)

I am interested in exploring the potential of digital planetariums and research grade visualization facilities. The term Data to Dome refers to a strategy to simplify and streamline the process of going from scientific discovery to planetarium visualization and presentation. This strategy included the development of new data and interoperability standards and well as the professional development of planetarium staff. We also hope to realize the power of the modern planetarium as a research grade visualization facility by introducing the capabilities in the software to go beyond presentation and enable interrogation of the data. With these new capabilities we have large collaborative data exploration and discovery facilities that have already been built and paid for in almost every major city of the world.

3.25 Interactive planetarium visualization

Edwin A. Valentijn (University of Groningen, NL)

We had extensive exchanges with the planetariums which are involved in interactive visualization and who are keen on bringing in new content on a very regular basis – also in relation to the Data-to-Dome initiative. Further progress of Data-to-Dome heavily depends on the involvement of the planetarium display system providers, Focus was a.o. on bringing current research into planetariums and to lower the threshold for researchers to present their work in a planetarium environment. A splinter session was organized with so called power users, Adler-Chicago, Morrison-San Francisco, Hayden – New York, Heidelberg, Norrköping, and DOTliveplanetarium–Groningen discussing common interests. In this context, the OpenSpace project was perceived as an important and promising way forward and new collaborations have been initiated. Real time streaming of 360 degree live content towards the planetariums was also discussed, but there was not a strong interest. Dotliveplanetarium is building a platform for this. In short, the workshop helped very much to settle new international relations and collaborations and helped to identify common interests. Some follow-up work visits have been planned.

In the seminar highlights we note: “We initiated to set up new collaborative IPS working groups for trans-national collaborations and exchange of data and content” Edwin Valentijn, Mark SubbaRao “We perceived a new and enthusiastic community dedicated to enhance planetarium visualisation for modern research, knowledge dissemination and outreach.”

3.26 Visualization, Relativity, and Astrographics

Daniel Weiskopf (Universität Stuttgart, DE)

My background is in visualization in general and covers a wide range of topics in scientific visualization, information visualization, and visual analytics. In particular, I am interested in the application of visualization related to astronomy and astrophysics. One example is visualization for special relativity [1, 4] and general relativity [5]. These kinds of visualization can be used for public outreach by incorporating concepts of explanatory and illustrative visualization, for example, for museum exhibitions [2]. Also, immersive virtual environments played a big role for such applications early on, as demonstrated for a virtual “flight simulator” for special relativity [3]. Another example of previous work in astrographics includes the reconstruction and rendering of nebulae, relying on efficient rendering methods and reconstruction schemes [6].

For future research, I am interested in further integrating astrographics and advanced visualization: How can astrographics benefit from recent developments in visualization research? And how will the special requirements of astrographics influence visualization? In particular, I see the need for further research on combining explanation and exploration in visualization, which directly links to questions related to methods for data-oriented storytelling. What are design principles and data-transformation methods that support storytelling? What are the differences between storytelling in astrographics and other application domains?

I also see a productive link between astrographics and other areas of visualization in the form of immersion: The recent trend toward immersive analytics brings 3D representations and immersion into the focus of visualization research [7]. How do these new approaches relate to already established ways of presenting visualizations in immersive planetarium environments? How can immersive analytics be used in astrographics?

References

1 Daniel Weiskopf. A Survey of Visualization Methods for Special Relativity. Scientific Visualization: Advanced Concepts, Dagstuhl Follow-Ups, Schloss Dagstuhl, p. 289–302, 2010
2 Daniel Weiskopf, Marc Borchers, Thomas Ertl, Martin Falk, Oliver Fechtig, Regine Frank, Frank Grave, Andreas King, Ute Kraus, Thomas Müller, Hans-Peter Nollert, Isabel Rica Mendez, Hanns Ruder, Tobias Schafhitzel, Sonja Schär, Corvin Zahn, Michael Zatloukal. Explanatory and Illustrative Visualization of Special and General Relativity. IEEE Transactions on Visualization and Computer Graphics, 12(4):522–534, 2006
3 Daniel Weiskopf. An Immersive Virtual Environment for Special Relativity. Proceedings of the International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG). p. 337–344, 2000
4 Daniel Weiskopf, Ute Kraus, Hanns Ruder. Searchlight and Doppler effects in the visualization of special relativity: a corrected derivation of the transformation of radiance. ACM Transactions on Graphics. 18(3):278–292, 1999
5 Daniel Weiskopf. Four-dimensional non-linear ray tracing as a visualization tool for gravitational physics. Proceedings of the IEEE Visualization 2000 Conference, p. 445–448, 2000
6 Stephan Wenger, Marco Ament, Stefan Guthe, Dirk A. Lorenz, Andreas M. Tillmann, Daniel Weiskopf, Marcus A. Magnor. Visualization of Astronomical Nebulae via Distributed Multi-GPU Compressed Sensing Tomography. IEEE Transactions on Visualization and Computer Graphics. 18(12):2188–2197, 2012
7 Kim Marriott, Falk Schreiber, Tim Dwyer, Karsten Klein, Nathalie Henry Riche, Takayuki Itoh, Wolfgang Stuerzlinger, Bruce H. Thomas. Immersive Analytics. Springer, 2018.

3.27 Astrographics in Context

Ryan Wyatt (California Academy of Sciences – San Francisco, US)

All visualizations require interpretation, and my interest as an astronomer turned educator center on leveraging visualizations for outreach to broad audiences. I am fundamentally a user of astrographics software, not a developer, so I want to consider how software can best enable educators to share visually compelling and scientifically accurate visualizations—and how these tools can support astronomical discovery.

I have several core interests in astrographics. First, the community must recognize the need to develop and refine presentation tools in astrographics software: particularly allowing for smooth transitions in scale (the ability to "fly" through data is key to providing continuity and context) and adjustability in terms of visual representations (essential for fine-tuning representations that will be accessible and comprehensible to a wide range of consumers). Second, our software needs to provide interpretive layers—both for educators and for our audiences—along the lines of the metadata enabled by the Astronomy Visualization Metadata (AVM) standard. Third, we need to make experiences interactive for our audiences, and although my primary interest is in mediated or guided experiences, it is also worth considering a broader definition that emphasizes agency on the part of novice users, although there are many attendant challenges—especially in terms of the scaffolding required for audiences to make sense of visualizations. Lastly, I am interested in ensuring that the wide range of tools (especially covering both research and education) remain extensible, compatible, and interoperable.

Dagstuhl Seminar 19262, "Astrographics: Interactive Data-Driven Journeys through Space," gave several of us the opportunity to consider how tools developed for educational use—specifically in planetarium domes—could support research. The resulting white paper, IDEAS: Immersive Dome Experiences for Accelerating Science, was submitted to the Astro2020 Decadal Survey on Astronomy and Astrophysics. As noted in the paper's abstract, "Planetariums can and should be used for the advancement of scientific research. [...] We propose a transformative model whereby scientists become the audience and explorers in planetariums, utilizing software for their own investigative purposes."

3.28 Exploranation – A New Science Communication Paradigm

Anders Ynnerman (Linköping University, SE)

The rapid development of computer hardware for visualization, together with new and improved methods and software, enables interactive exploration of complex and large-scale data on commodity graphics processing units (GPUs). Scientific data is also to an increasing degree published as "Open Data". These three trends together make it possible to let visitors to public venues, such as science centers and museums, themselves become explorers of scientific data using the same tools, methods, data and even hardware that researchers are using in scientific exploration. This change of use of visualization in public spaces from the traditional explanatory visualization to an exploratory approach has been referred to as "Exploranation", a euphemism encapsulating the confluence of explanation and exploration.

Exploranation, however, is in need of further research as challenges appear when interactive visualization is used in science communication. For installations enabling free exploration intuitive and robust interaction are key features. It is also important to address issues arising in non-linear storytelling. If there are specific learning goals expressed, non-invasive guidance has to be used to ensure that goals are reached and this without disturbing the notion of free exploration.

In the case of live facilitated presentations using interactive data exploration many of these challenges also appear. In addition, the role and importance of the facilitator needs to be further researched in the context of performativity and the expression and transfer of tacit knowledge. An important aspect of live presentations is also the navigation and piloting. To enable a facilitator/presenter to both navigate and present, high level tools combining story telling with navigation and data processing need to be developed.

4 Break-out Groups

4.1 Domecasting + Authoring Tools

Alexander Bock (Linköping University, SE & University of Utah, US), Brian Abbott (AMNH – New York, US), Charles D. Hansen (University of Utah – Salt Lake City, US), and Wolfgang Steffen (Universidad Nacional Autonoma de Mexico – México, MX)

This breakout session focussed around the use of astrovisualization tools in the context of domecasting, and how to create authoring tools such that novice users are able to quickly generate new content that can be shared with a wide audience. Domecasting in this context refers to the ability to connect multiple instances of a software together and replicate the input from one master computer to connected clients. This technique can, for example be used to share a live presentation given by a subject-matter expert in a specific location and share both audio and visual results with multiple other locations that are located in other places around the globe. This method relies on the ability for each client installation to run the visualization tool locally such that images can be generated which are optimal for the geometry of the facility. This is necessary as the connected geometries might vary greatly, from flat screens, to virtual reality headsets, and planetariums, making a custom solution for each installation an essential requirement. While discussing the technical details of the currently implemented solution of open-source software in this area, it was determined that each vendor/provider uses a closed system that are currently unable to interoperate. This lead to the discussion to create a common format that can be used to allow domecasting between different astrovisualization software vendors and provide a common interface that new software can integrate. While this method does not ensure that all vendors will provide the same functionality, it is the first step in ensuring that these features can be implemented in the future. A substantial part of this discussion was also devoted to the concept of recording and authoring flight paths, i.e. the path of the virtual camera in both space and time through the visualization. Representations of these flight paths might be standardized to the extent that they can be shared between different vendors. Furthermore, participants voiced their wish for improving the current state of these flight paths to make it possible to add additional information to these flight paths that is shown to a potential presented in a public environment, informing them about potential topics that might be discussed while automatically following a set camera movements. This thread of sparked a lengthy discussion on a concept called “Showkits” which were physical materials and props prepared by the American Museum of Natural History many years ago that could be sent to school teachers and other museums to support their efforts while giving a live presentation. Annotated flight paths that also include additional media would be moving this concept to the modern age and would make it possible to disseminate the didactics on how to give live presentations to a larger audience, and would also provide the ability to create a central location to share these showkits to train new pilots or have people self-train based on these materials. The end result of this part of the discussion was defining terminology for future work surrounding this concept. A modern "Showkit" is a directed graph of "showlets" which can be navigated by the user at their own pace. A "showlet" is a collection of assets including media resources, such as audio transcripts of a subject-matter expert, videos to be included in the presentation, and presenter notes for the pilot of the presentation. The connection between different "showlets" would occur through the use of either pre-recorded or dynamically generated flight paths. A similarly lively discussion occurred around the topic of authoring tools that would provide a larger amount of people with the possibility to create new content that can be shared with a large number of planetariums. First and foremost on the agenda was the topic of interoperability between different vendors. A perfect content authoring tool would export the resulting content in a format that can be easily converted between different vendors such that content only needs to be generated once and can be reused by any vendor. A long in-depth discussion followed on the merits of providing different complexities of authoring tools, one for the casual user that does not provide as many detailed options, all the way to the expert user who can include specialized code for rendering new datasets, which can then be shared with other users. Returning the discussion to the showlets, it was determined that it would be most beneficial to have different tools to be able to create individual assets, then combining different assets into showlets using a separate application and then finally integrating different showlets into entire shows or showkits using yet another application. Using this three-stage process it would be possible to specialize each tool for its intended purpose and optimize it for its niche.

4.2 Quo vadis Astrovisualization software?

Alexander Bock (Linköping University, SE & University of Utah, US), Tom Kwasnitschka (GEOMAR Helmholtz Centre for Ocean Research Kiel, DE), David H. Laidlaw (Brown University – Providence, US), and Thomas Müller (Haus der Astronomie – Heidelberg, DE)

This break-out session investigated where different publicly available astrovisualization software should go and focus their efforts on to be effective tools at communicating astronomy and astrophysics research. These discussions were not each explored in detail, but it was rather designed as a more general brainstorming sessions. The first large component of desired features included support for GIS (Geographic information system) algorithms that would make it possible for users to import their own specialized data. One aspect of this is having the ability for scientists to add custom data loaders for their own specialized formats to shorten the path between exporting data from specialized analysis tools into the visualization frameworks. These also include supporting new rendering primitives, such as raster images, polygons, and geolocators, but also reducing the friction to import other external data sources, such as georeferenced images in common data formats, such as GeoTIFF. Desired operations for these GIS datasets are mostly related to measuring; being able to measure height information, the value of individual pixels, and other tools to verify the visualization. A concrete example for these tools are the ocean/bay science, where information for most of the Earth's ocean surface is available and has not yet been utilized for visualization purposes or public outreach. Another wish for astrovisualization tools is the increased support for more virtual reality devices. These include virtual reality headsets, both commercially available headsets as well as research devices, but also CAVE environments.

Both directions require further research before they can be used in regular programming for a wider audience as, for example, guided navigation techniques have to be further developed first. Another type of large data that is currently not handled broadly is point cloud data, particularly data acquired from LiDAR sensors. Participants were particularly focussed on the ocean surface scanning that has occurred over the past years, which resulted in a large amount of available point cloud data that is currently underutilized due to the lack of visualization techniques. Issues that were raised are the handling of memory requirements as these datasets usually tend to be very large, and special care needs to be taken when simplifying them. One suggestion simplification technique that was discussed in detail were methods of transforming the LiDAR point clouds into polygonal representations that could be displayed and stored much more efficiently and would not lose much information in the process. Lastly, the participants discussed potentially adding complex rendering techniques to these softwares, such as wormhole or blackhole renderings that would operate in real-time. These techniques would make it possible to illustrate the effects of gravity for high-school level students that already have some familiarity with the matter, but lack intuitive representations except for high-budget hollywood movies. Providing these students with the ability to change gravitational parameters and allowing them to dynamically inspect the results could potentially have large benefits in conveying these complex topics.

4.3 Towards Augmented Reality Astronomy Research in the mid 2020s

Melvyn Davies (Lund Observatory, SE), Dave Brown (Microsoft Research – Redmond, US), Alyssa A. Goodman (Harvard-Smithsonian Center for Astrophysics, US), and Joshua Eli Goldston Peek (Space Telescope Science Inst. – Baltimore, US)

Much of astronomy research requires collaborative, high-dimensional data exploration. It is collaborative because astronomers measure the same objects and structures with many different methods, and no one person can be an expert in all technical domains. It is high-dimensional because each of these methods can add many new dimensions to each object. Software like glue can allow a single user to explore these data, but when we are talking about large complex data sets in multiple dimensions, the immediacy and structure of true 3D representations in real space dramatically increases the depth of understanding of relative scale and structure.

Virtual reality offers a new window into this world, with much advanced tools developed by VR experts and powered by the gaming industry. But VR structurally limits collaboration – communication between astronomers happens via facial interaction and body posture, which is why most remote group meetings happen via video chat, rather than simply audio call.

Our group investigated the idea of how augmented reality and mixed reality might open up a space for collaborations of research astronomers to explore data together. We used a mixed-reality headset (a version 1 HoloLens) to drive home the point that data exploration is much richer in 3D.

VR experts explained the various pros and cons of VR headsets and tooling; how progress has been made in making the tools better for data exploration. The concept of the ease of access curve, with viz-cave most difficult to access, shading to VR and MR devices, all the way to AR-enabled phones, which most participants already had, was discussed. We focused on practical matters – what it would look like for research astronomers to mix in AR and MR tooling into their daily workflow.

The concept of merging glue, HoloLens, and ARKit2/Tango (phone) AR technologies was discussed. It was agreed that such an arrangement made some sense as a toolchain for casual visual exploration and was not out of reach – glue can stage the data and cast it to the internet, a single MR user could see a 3D data structure and provide 3D brushing capabilities, and anyone with a phone could participate in viewing the 3D space.

4.4 Decadal White Paper: Immersive Dome Experiences for Accelerating Science

Jackie Faherty (AMNH – New York, US), Brian Abbott (AMNH – New York, US), Alexander Bock (Linköping University, SE & University of Utah, US), Dave Brown (Microsoft Research – Redmond, US), David H. Laidlaw (Brown University – Providence, US), Marcus A. Magnor (TU Braunschweig, DE), Thomas Müller (Haus der Astronomie – Heidelberg, DE), Joshua Eli Goldston Peek (Space Telescope Science Inst. – Baltimore, US), Wolfgang Steffen (Universidad Nacional Autonoma de Mexico – México, MX), Gabriel Stöckle (Naturhistorisches Museum Wien, AT), Mark Subbarao (Adler Planetarium – -Chicago, US), Daniel Weiskopf (Universität Stuttgart, DE), Ryan Wyatt (California Academy of Sciences – San Francisco, US), and Anders Ynnerman (Linköping University, SE)

Astrophysics lies at the crossroads of big datasets (such as the Large Synoptic Survey Telescope and Gaia), open source software to visualize and interpret high dimensional datasets (such as Glue, WorldWide Telescope, and OpenSpace), and uniquely skilled software engineers who bridge data science and research fields. At the same time, more than 4,000 planetariums across the globe immerse millions of visitors in scientific data. We have identified the potential for critical synergy across data, software, hardware, locations, and content that—if prioritized over the next decade—will drive discovery in astronomical research. Planetariums can and should be used for the advancement of scientific research. Current facilities such as the Hayden Planetarium in New York City, Adler Planetarium in Chicago, Morrison Planetarium in San Francisco, the Iziko Planetarium & Digital Dome Research Consortium in Cape Town, and Visualization Center C in Norrköping are already developing software which ingests catalogs of astronomical and multi-disciplinary data critical for exploration research primarily for the purpose of creating scientific storylines for the general public. We propose a transformative model whereby scientists become the audience and explorers in planetariums, utilizing software for their own investigative purposes. In this manner, research benefits from the authentic and unique experience of data immersion contained in an environment bathed in context and equipped for collaboration. Consequently, in this white paper we argue that over the next decade the research astronomy community should partner with planetariums to create visualization-based research opportunities for the field. Realizing this vision will require new investments in software and human capital.

4.5 Presentation & Authoring

Jackie Faherty (AMNH – New York, US), Brian Abbott (AMNH – New York, US), Hidehiko Agata (NAOJ – Tokyo, JP), Alexander Bock (Linköping University, SE & University of Utah, US), Dave Brown (Microsoft Research – Redmond, US), Charles D. Hansen (University of Utah – Salt Lake City, US), Tom Kwasnitschka (GEOMAR Helmholtz Centre for Ocean Research Kiel, DE), Wolfgang Steffen (Universidad Nacional Autonoma de Mexico – México, MX), Gabriel Stöckle (Naturhistorisches Museum Wien, AT), Mark Subbarao (Adler Planetarium – Chicago, US), Ryan Wyatt (California Academy of Sciences – San Francisco, US), and Anders Ynnerman (Linköping University, SE)

Astrographics come to life when coupled with story—either in a live presentation or a pre-recorded narrative. Presentation and authoring are thus core functions of astrographics tools, and we address priorities and goals for improving these functions.

As one participant noted, "personal presentation is the most powerful means of communication," and live presentation forms the core of the astrographics experience in planetariums and other venues. The community should support presenters in becoming the best they can be, and both training and tool development were identified as key in advancing the field.

There is also the desire to capture the essential qualities of a live presentation—for sharing with other presenters and venues, for domecasting from one theater to another, or for authoring a playback version of the content

4.6 Software Integration

Alyssa A. Goodman (Harvard-Smithsonian Center for Astrophysics, US), Emil Axelsson (Linköping University, SE), Thomas Boch (Université de Strasbourg – Strasbourg, FR), Carter Emmart (AMNH – New York, US), Andreas Gerndt (DLR – Braunschweig, DE), David H. Laidlaw (Brown University – Providence, US), Thomas Müller (Haus der Astronomie – Heidelberg, DE), Thomas P. Robitaille (Aperio Software – Leeds, GB), and Edwin A. Valentijn (University of Groningen, NL)

It can be argued that before this seminar that the planetarium was not a scientific tool. Merging the tools and methods of science into the immersive dome display environment of the planetarium has been envisioned but the results of this gathering have enabled the fledgling steps toward this goal. With capacities upward to several hundred in one place, and ability to network such facilities together worldwide the planetarium community has always held the potential for wide scale group immersion. From this point forward, scientists with tools familiar to them can now use the global planetarium network for focused interaction with their data to concentrate and multiple mental solutions on vital scientific line of inquiry. While tested in academia and industry it can be argued that the results of this seminar have enabled a vital linkage for large scale data immersion like never before, which in essence is a paradigm shift.

4.7 How do we, planetaria, get better together and share / stream content?

Edwin A. Valentijn (University of Groningen, NL), Carter Emmart (AMNH – New York, US), Jackie Faherty (AMNH – New York, US), Mark Subbarao (Adler Planetarium – Chicago, US), and Ryan Wyatt (California Academy of Sciences – San Francisco, US)

We had an intense brainstorming session in which we have assessed many current aspects of collaborations in between planetaria and with vendors. We summarize the recommendations phrased at this brainstorm on various topics.

- Streaming in real time full dome live content from an event to dome(s): There is not not so much interest in the general community, but DOTliveplanetarium will pursue this and report on its merits to the other participants.
- Streaming in real time live full dome content between different planetariums (domecasting): This is an interesting mode of operation and allows sharing content. OpenSpace can/should be the preferred platform for this, facilitating both sending and receiving streams. We suggest a community effort to help OpenSpace to achieve this goal.

4.8 Support for Analysis Tools

Anders Ynnerman (Linköping University, SE)

This breakout session focussed on aspects of data analysis. The broad-ranging discussion can be grouped into three topics: flow from data discovery to data analysis; making the visualization traceable; and novel analysis techniques with a particular focus on Artificial Intelligence/Machine Learning.

A challenge in data discovery is simply to be aware of a certain data set. We discussed how one might help people find data sets. We also noted how as a starting astronomer, it's hard to combine things. Data integration is elusive as compared to working within a single data set. Data sources and links embedded in images have a lot to do with lineage and traceability, and the smooth transition from discovery to analysis. Visual flow is needed for really low friction interactions (find the data quickly) – tools of visualization can be used for data discovery before exploration. Broad explorations needed for hypothesis generation. In our view, we haven't yet applied the methods of visualization to questions of data discovery

How does one best capture a visualization session? Being able to reproduce the dialogue of such a session has benefits in many domains: it can help in teaching and training users and the promotion of best practice. Done well, it can aid in scientific discovery. We discussed a number of issues and technical details related to traceability. It is not obvious how best to capture the visualization. For example in going from the initial data to the final visualization, one may take a relatively circuitous path with a number of blind alleys. Does one remove some of these? All of these? Mistakes can provide valuable lessons. They may lead to discoveries of ways forward by other subsequent users. One should therefore compress with caution. We have code for tracking code (version control). Do we have visualization for tracking visualization? We noted in the discussion that viztrails goes somewhat in this direction.

Very much connected with the above discussion are the prospects to use AI/ML to aid users. This support can take many formats. For example, recommender systems can help users to find data sets (the unknown unknowns) – those who looked at X, also looked at Y…We noted how this could be applied to both data sets and research papers. In a broad sense, we identified three types of AI augmentation: one could help speed the visualization process along by doing the same functions as a user but just more quickly; one could also provide faster fitting; and perhaps most significantly one could provide deeper insight. In this third type of support, we envisage the AI system to help show the way, ie help turn a novice user into a more-experienced user. We did however note some possible pitfalls. The AI could for example make erroneous inferences about the users intentions: for example recommending colour maps to demarcate certain parts of the data whereas in reality the user is interested in other parts: emphasising the peaks when the user wishes to find the troughs.

4.9 Virtual Presenters

Anders Ynnerman (Linköping University, SE) and Carter Emmart (AMNH – New York, US)

An exciting new development in computer graphics and data analysis is the embodiment of the visualization/narration system itself as an avatar, or cognitive companion. For some tasks, such as training, natural interaction, and presentation of aggregated information, the avatar could be manifested as a photo-real digital human. For other tasks, the cognitive companion can be represented as visualizations of multi-source aggregated data. One of the most natural human-friendly approaches to interfacing to the system is for the user to simply be able to ask informed questions about the application domain, and to expect contextualized, insightful answers. For instance, whether the modality used in the interface is an avatar or visualization of complex data, one still wants to ask questions to the avatar, or about the visualization, as if one is having a dialogue with a literate, expert, in short a cognitive assistant.

As an example of possibilities for new levels of interaction with human representations are rapidly opening up through recent advances in visual representation of human avatars imaging using light field displays supporting virtual human style interaction. A light field display is a system which offers viewing of stereoscopic 3D-images from arbitrary positions without 3D-glasses. Compared to traditional 2D displays which shows a 2D image these, new displays show 4D light fields, which for each pixel encodes the angular intensity variations making it a see-through window into the virtual 3D-world. Light field viewing experiences will fundamentally change our notion of a display, how we use it and how we develop content for it.

An exciting possibility is thus the development of a Q&A systems tailored for visual science communication. IBM introduced Watson (www.ibm.com), a technology platform based on natural language processing and machine learning intended to analyze large amounts of unstructured data, and to query the results to provide deep insight about a particular application domain. The combination of natural interaction with human representations using technologies such as light field displays, where avatar representations act as conveyors of information and ultimately, with such a Q&A systems could lead to a virtual guide that would further increase the possibilities to provide engaging and explainable visualization for general audiences.

Hand-held navigation interface for controlling dome software via a smartphone or similar pressure-sensitive touchscreen device would be useful. In the planetarium setting, a presenter typically discusses with the audience while the pilot must remain at a desktop computer station to drive and navigate the presentation's software. For even simple programs to be engaging, a production necessitates two performers: the pilot and a presenter, who can remain visible and accessible to the audience. Similar to a drone controller, a handheld device to control the camera path would eliminate the pilot. Having virtual joystick controls for the left/right thumbs allows arbitrary (within the constraints of dome comfort) camera motions for flying through a scene. Additionally, one could use the accelerometers of a device such as an iPhone to control the camera path freeing the presenter to focus on the dome and move through a scene.

Participants

- Brian Abbott
AMNH – New York , US
- Hidehiko Agata
NAOJ – Tokyo, JP
- Emil Axelsson
Linköping University, SE
- Thomas Boch
Université de Strasbourg – Strasbourg, FR
- Alexander Bock
Linköping University, SE & University of Utah, US
- Dave Brown
Microsoft Research – Redmond, US
- Melvyn Davies
Lund Observatory, SE
- Carter Emmart
AMNH – New York, US
- Jackie Faherty
AMNH – New York, US
- Andreas Gerndt
DLR – Braunschweig, DE
- Alyssa A. Goodman
Harvard-Smithsonian Center for Astrophysics, US
- Charles D. Hansen
University of Utah – Salt Lake City, US
- Tom Kwasnitschka
GEOMAR Helmholtz Centre for Ocean Research Kiel, DE
- David H. Laidlaw
Brown University – Providence, US
- Marcus A. Magnor
TU Braunschweig, DE
- Thomas Müller
Haus der Astronomie – Heidelberg, DE
- Joshua Eli Goldston Peek
Space Telescope Science Inst. – Baltimore, US
- Lucian Plesea
ESRI – Redlands, US
- Sebastian Ratzenböck
Universität Wien, AT
- Thomas P. Robitaille
Aperio Software – Leeds, GB
- Filip Sadlo
Universität Heidelberg, DE
- Wolfgang Steffen
Universidad Nacional Autonoma de Mexico – México, MX
- Gabriel Stöckle
Naturhistorisches Museum Wien, AT
- Mark Subbarao
Adler Planetarium – Chicago, US
- Edwin A. Valentijn
University of Groningen, NL
- Daniel Weiskopf
Universität Stuttgart, DE
- Ryan Wyatt
California Academy of Sciences – San Francisco, US
- Anders Ynnerman
Linköping University, SE

Report from Dagstuhl Seminar 19271

Graph Colouring: from Structure to Algorithms

Edited by

Maria Chudnovsky[1], Daniel Paulusma[2], and Oliver Schaudt[3]

1 **Princeton University, US,** mchudnov@math.princeton.edu
2 **Durham University, GB,** daniel.paulusma@durham.ac.uk
3 **RWTH Aachen, DE,** schaudt@mathc.rwth-aachen.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 19271 "Graph Colouring: from Structure to Algorithm", which was held from 30 June to 5 July 2019. The report contains abstracts for presentations about recent structural and algorithmic developments for the Graph Colouring problem and variants of it. It also contains a collection of open problems on graph colouring which were posed during the seminar.

Seminar June 30–July 5, 2019 – http://www.dagstuhl.de/19271
2012 ACM Subject Classification Mathematics of computing → Graph coloring, Mathematics of computing → Graph algorithms
Keywords and phrases (certifying / parameterized / polynomial-time) algorithms, computational complexity, graph colouring, hereditary graph classes
Digital Object Identifier 10.4230/DagRep.9.6.125
Edited in cooperation with Cemil Dibek

1 Executive Summary

Maria Chudnovsky (Princeton University, US)
Daniel Paulusma (Durham University, GB)
Oliver Schaudt (RWTH Aachen, DE)

The Graph Colouring problem is to label the vertices of a graph with the smallest possible number of colours in such a way that no two neighbouring vertices are identically coloured. Graph Colouring has been extensively studied in Computer Science and Mathematics due to its many application areas crossing disciplinary boundaries. Well-known applications of Graph Colouring include map colouring, job or timetable scheduling, register allocation, colliding data or traffic streams, frequency assignment and pattern matching. However, Graph Colouring is known to be computationally hard even if the number of available colours is limited to 3.

The central research aim of our seminar was to increase our understanding of the computational complexity of the Graph Colouring problem and related NP-complete colouring problems, such as Precolouring Extension, List Colouring and H-Colouring. The approach followed at the seminar for achieving this aim was to restrict the input of a colouring problem to some special graph class and to determine wether such a restriction could make the problem tractable.

As input restriction, the main focus was to consider hereditary graph classes, which are those classes of graphs that are closed under vertex deletion. Hereditary graph classes provide a unified framework for a large collection of well-known graph classes. The reason for this is that a graph class is hereditary if and only if it can be characterized by a (unique) set $\mathcal{H}$ of minimal forbidden induced subgraphs. This property enables a *systematic* study into the computational complexity of a graph problem under input restrictions. For instance, one can first restrict the input to some hereditary graph class for which $\mathcal{H}$ is small, say $\mathcal{H}$ has size 1 or 2, or for which $\mathcal{H}$ consists of small graphs only.

In line with the seminar's research aim, the seminar brought together researchers from Discrete Mathematics, working in structural graph theory, and researchers from Theoretical Computer Science, working in algorithmic graph theory. In total, 45 participants participated from 14 different countries.

The scientific program of the seminar consisted of 23 sessions: 4 one-hour survey talks, 17 contributed talks of at most thirty minutes and 2 open problem sessions. This left ample time for discussions and problem solving.

Each of the four survey talks covered a particular structural or algorithmic key aspect of the seminar to enable collaborations of researchers with different backgrounds. On Monday, Sophie Sprikl presented a state-of–the-art summary of the Graph Colouring problem for H-free graphs and gave the main ideas and techniques behind an important, recent result in the area, namely a polynomial-time algorithm for colouring P_6-free graphs with at most four colours. On Tuesday, Marcin Pilipczuk gave a tutorial on the framework of minimal chordal completions and potential maximal cliques. This technique plays a crucial role for solving the Maximum Independent Set problem on some hereditary graph classes, but has a much wider applicability. On Wednesday, Bart Jansen gave a presentation on the parameterized complexity of the Graph Colouring problem and related colouring problems. Due to a large variety of possible parameterizatons, Jansen's talk covered a wide range of open problems. On Thursday, Konrad Dabrowski gave an introduction to the clique-width of hereditary graph classes. If a graph class has bounded clique-width, then Graph Colouring and many other NP-hard problems become polynomial-time solvable. Hence, as a first step in the design of a polynomial-time algorithm, one may first want to verify if the clique-width (or any equivalent width parameter) of the graph class under consideration is bounded.

The two general open problem sessions took place on Monday and Tuesday afternoon. Details of the presented problems can be found in the report, together with abstracts of all the talks.

2 Table of Contents

3 Overview of Talks

3.1 Revisiting a theorem by Folkman on graph colouring

Marthe Bonamy (University of Bordeaux, FR)

Joint work of Marthe Bonamy, Pierre Charbit, Oscar Defrain, Gwenaël Joret, Aurélie Lagoutte, Vincent Limouzy, Lucas Pastor, Jean-Sébastien Sereni
Main reference Marthe Bonamy, Pierre Charbit, Oscar Defrain, Gwenaël Joret, Aurélie Lagoutte, Vincent Limouzy, Lucas Pastor, Jean-Sébastien Sereni: "Revisiting a theorem by Folkman on graph colouring", CoRR, Vol. arXiv:1907.11429, 2019.
URL https://arxiv.org/abs/1907.11429v1

We give a short proof of the following theorem due to Jon H. Folkman [1]: The chromatic number of any graph is at most 2 plus the maximumover all sub-graphs of the difference between half the number of vertices and the independencenumber.

References

1 Folkman, J. H.. *An upper bound on the chromatic number of a graph* (No. RM-5808-PR). RAND CORP SANTA MONICA CALIF, 1969.

3.2 On an augmenting graph approach for the maximum-weight independent set problem

Christoph Brause (TU Bergakademie Freiberg, DE)

The augmenting graph technique is an approach that solves the maximum independent set problem in various graph classes polynomially. Although we know a little about this technique, our knowledge about implementations for the maximum-weight independent set problem is very limited.

In this talk, we present a polynomial-time augmenting graph approach for the weighted version and some suitable graph classes, e.g. subclasses of $S_{1,k,k}$-free graphs, and consider a combination with decompositions by clique separators.

3.3 Introduction to Clique-width and Open Problems

Konrad Dabrowski (Durham University, GB)

Joint work of Konrad K. Dabrowski, Matthew Johnson, Daniël Paulusma
Main reference Konrad K. Dabrowski, Matthew Johnson, Daniël Paulusma: "Clique-Width for Hereditary Graph Classes," London Mathematical Society Lecture Note Series 456:1-56, Cambridge University Press, 2019.
URL https://doi.org/10.1017/9781108649094.002

Graphs classes of bounded clique-width are interesting from a computational perspective, because many problems, such as Colouring, are polynomial-time solvable on such classes. I will give an introduction to clique-width and explain some of the techniques at our disposal when dealing with this parameter. I will also present a number of open problems on boundedness of clique-width for various graph classes and some related problems on Colouring. See also our survey https://arxiv.org/abs/1901.00335

3.4 Coloring graphs by forbidden induced subgraphs

Chinh T. Hoàng (Wilfrid Laurier University – Waterloo, CA)

Joint work of Yingjun Dai, Dallas J. Fraser, Angèle M. Hamel, Chính T. Hoàng, Frédéric Maffray
Main reference Dallas J. Fraser, Angèle M. Hamel, Chính T. Hoàng, Frédéric Maffray: "A coloring algorithm for -free line graphs", Discrete Applied Mathematics, Vol. 234, pp. 76–85, 2018.
URL https://doi.org/10.1016/j.dam.2017.06.006

Let F_4 be a set of four-vertex graphs. For any set F_4, it is known that COLORING F_4-free graphs is NP-hard or solvable in polynomial time, except when F_4 is one of the following three sets: {claw, $4K_1$}, {claw, $4K_1$, co-diamond}, $\{4K_1, C_4\}$. In this talk, we survey recent advances on these three open problems. We will discuss the two tools that have been proved to be useful in attacking the problems: perfect graph theory, and the theory of clique width.

3.5 Shitov's Counterexample to Hedetniemi Conjecture

Shenwei Huang (Nankai University – Tianjin, CN)

Hedetniemi conjecture is a well-known conjecture in the study of graph coloring. It remained open for 53 years until two months ago Shitov came up with a counterexample and hence disproved the conjecture. The proof is basic but elegant. In this talk, we will present the proof of Shitov's counterexample.

3.6 Parameterized Complexity of Graph Coloring Problems

Bart Jansen (TU Eindhoven, NL)

This talks surveys various aspects of the parameterized complexity of graph coloring problems. The goal is to understand how certain complexity parameters contribute to the difficulty of finding exact solutions to such problems. We discuss results in various parameterized algorithmic regimes, and point out open problems wherever possible. The regimes we consider are:

- Fixed-parameter tractable algorithms, for parameterizations that capture the structural complexity of the input graph. We will look at questions such as: if graph G is only k vertex deletions away from belonging to a graph class where coloring is easy, then can the coloring problem on G by solved in $f(k)n^c$ time for some function f and constant c?
- Fixed-parameter tractable algorithms that work on a decomposition of the input graph. Given a graph G and a tree decomposition of width w, one can test the q-colorability of G in time $O^*(q^w)$, which is essentially optimal assuming the Strong Exponential Time Hypothesis. We will see how working over a linear layout of cutwidth w allows the problem to be solved much faster, by exploiting an interesting connection to the rank a matrix that encodes the compatibility of colorings on two sides of small edge cut.

- Fixed-parameter tractable algorithms for parameterizations that measure how far the input graph violates conditions that guarantee the existence of a good coloring. Brooks' theorem guarantees that any graph G that is not a clique or odd cycle, can be colored with $\Delta(G)$ colors. Hence it is easy to test if a graph whose vertices have degree at most q, can be q-colored. How hard is it to test if G has a coloring with q colors, when only k vertices of G have degree more than q?
- Kernelization algorithms. Let k be a parameter that captures the structural complexity of the input graph – for example, the size of a minimum vertex cover. Is it possible to preprocess an input G in polynomial time, obtaining a graph G' of size polynomial in k, so that G has a 3-coloring if and only if G' has one? What is the best upper-bound on the size of G' in terms of k?

3.7 Classes with no long cycle as a vertex-minor are polynomially chi-bounded

O-joung Kwon (Incheon National University, KR)

Joint work of O-joung Kwon, Ringi Kim, Sang-Il Oum, Vaidy Sivaraman

A class $\mathcal{G}$ of graphs is χ-bounded if there is a function f such that for every graph $G \in \mathcal{G}$ and every induced subgraph H of G, $\chi(H) \leq f(\omega(H))$. In addition, we say that $\mathcal{G}$ is *polynomially χ-bounded* if f can be taken as a polynomial function. We prove that for every integer $n \geq 3$, there exists a polynomial f such that $\chi(G) \leq f(\omega(G))$ for all graphs with no vertex-minor isomorphic to the cycle graph C_n. To prove this, we show that if $\mathcal{G}$ is polynomially χ-bounded, then so is the closure of $\mathcal{G}$ under taking the 1-join operation.

3.8 The size Ramsey number of graphs with bounded treewidth

Anita Liebenau (UNSW Sydney, AU)

Joint work of Anita Liebanau, Nina Kamcev, David R Wood, Liana Yepremyan
Main reference Nina Kamcev, Anita Liebenau, David Wood, Liana Yepremyan: "The size Ramsey number of graphs with bounded treewidth", CoRR, Vol. abs/1906.09185, 2019.
URL http://arxiv.org/abs/1906.09185

A graph G is Ramsey for a graph H if every 2-colouring of the edges of G contains a monochromatic copy of H (not necessarily induced). The size Ramsey number of H is the smallest number of edges of a graph G that is Ramsey for H. This parameter received a lot of attention, in particular for sparse graphs H. We generalise earlier work and show that if the maximum degree and treewidth of H are bounded, then the size Ramsey number is linear in $|V(H)|$.

3.9 Cliquewidth III: The Odd Case of Graph ColoringParameterized by Cliquewidth

Daniel Lokshtanov (University of California – Santa Barbara, US)

Joint work of Fedor V. Fomin, Petr A. Golovach, Daniel Lokshtanov, Saket Saurabh, Meirav Zehaviv
Main reference Fedor V. Fomin, Petr A. Golovach, Daniel Lokshtanov, Saket Saurabh, Meirav Zehavi: "Clique-width III: Hamiltonian Cycle and the Odd Case of Graph Coloring", ACM Trans. Algorithms, Vol. 15(1), pp. 9:1–9:27, 2019.
URL https://doi.org/10.1145/3280824

Max-Cut(MC), Edge Dominating Set(EDS), Graph Coloring(GC) and Hamiltonian Path(HP) on graphs of bounded cliquewidth have received significant attention, as they can be formulated in MSO2 (and therefore have linear-time algorithms on bounded treewidth graphs by the celebrated Courcelle's theorem), but cannot be formulated in MSO1 (which would have yielded linear-time algorithms on bounded cliquewidth graphs by a well-known theorem of Courcelle, Makowsky, and Rotics). Each of these problems can be solved in time $g(k)\ n^{f(k)}$ on graphs of cliquewidth k. Fomin et al. [Intractability of Clique-Width Parameterizations. SIAM J. Comput. 39(5): 1941-1956 (2010)] showed that the running times cannot be improved to $g(k)n^{O(1)}$ assuming $W[1] \neq FPT$. However, this does not rule out non-trivial improvements to the exponent $f(k)$ in the running times. In a follow-up paper, Fomin et al. [Almost Optimal Lower Bounds for Problems Parameterized by Clique-Width. SIAM J. Comput. 43(5): 1541-1563 (2014)] improved the running times for EDS and MC to $n^{O(k)}$, and proved $g(k)n^{o(k)}$ lower bounds for EDS, MC and HP assuming the ETH. Recently, Bergougnoux, Kante and Kwon [WADS 2017] gave an $n^{O(k)}$-time algorithm for HP. Thus, prior to this work, EDS, MC and HP were known to have tight $n^{\Theta(k)}$ algorithmic upper and lower bounds. In contrast, GC has an upper bound of $n^{O(2^k)}$ and a lower bound of merely $n^{o(k^{1/4})}$ (implicit from the W[1]-hardness proof). Here we close the gap for GC by proving a lower bound of $n^{2^{o(k)}}$. This shows that GC behaves qualitatively different from the other three problems. To the best of our knowledge, GC is the first natural problem known to require exponential dependence on the parameter in the exponent of n.

3.10 Flexibility of Planar Graphs

Tomáš Masařík (Charles University – Prague, CZ)

Joint work of Zdenek Dvorák, Tomás Masarík, Jan Musílek, Ondrej Pangrác
Main reference Zdenek Dvorák, Tomás Masarík, Jan Musílek, Ondrej Pangrác: "Flexibility of triangle-free planar graphs", CoRR, Vol. abs/1902.02971, 2019.
URL https://arxiv.org/abs/1902.02971

Proper graph coloring assigns different colors to adjacent vertices of the graph. Usually, the number of colors is fixed or as small as possible. Consider applications (e.g. variants of scheduling) where colors represent limited resources and graph represents conflicts, i.e., two adjacent vertices cannot obtain the same resource. In such applications, it is common that some vertices have preferred resource(s). However, unfortunately, it is not usually possible to satisfy all such preferences. The notion called flexibility was recently defined by Dvořák, Norin, and Postle [1]. There instead of satisfying all the preferences the aim is to satisfy at least a constant fraction of any request.

We introduce main technical tools in the area and we present a structural statement for:

- Planar graphs without 4-cycles and with lists of size at least five [4].
- Planar graphs without triangles and with lists of size at least four [3].
- Planar graphs of girth at least 6 and with lists of size at least three [2].

We derive the following statement for all of them. Let G be an above-defined graph with a list assignment L. There exists an absolute constant such that for any (weighted) choice of preferred colors for some of the vertices, there is an L-coloring respecting at least a constant fraction of the preferences.

References

1 Z. Dvořák, S. Norin, and L. Postle, *List coloring with requests*, Journal of Graph Theory, (2019).

2 Z. Dvořák, T. Masařík, J. Musílek, and O. Pangrác, *Flexibility of planar graphs of girth at least six*, arXiv:1902.04069 (2019).

3 Z. Dvořák, T. Masařík, J. Musílek, and O. Pangrác, *Flexibility of triangle-free planar graphs*, arXiv:1902.02971 (2019).

4 T. Masařík, *Flexibility of planar graphs without 4-cycles*, Accepted to Eurocomb 2019, arXiv:1903.01460 (2019).

3.11 The Erdős-Hajnal property for graphs with no fixed cycle as a pivot-minor

Sang-il Oum (IBS – Daejeon, KR)

Joint work of Sang-il Oum, Jaehoon Kim

We prove that for every integer k, there exists $\varepsilon > 0$ such that every n-vertex graph with no pivot-minors isomorphic to C_k, the cycle graph on k vertices, has a pair of disjoint sets A, B of vertices such that $|A|, |B| \geq \varepsilon n$ and A is complete or anticomplete to B. This proves the analog of the Erdős-Hajnal conjecture for the class of graphs with no pivot-minors isomorphic to C_k.

3.12 Computing the chromatic number of a ring

Irena Penev (Charles University – Prague, CZ)

Joint work of Frédéric Maffray, Irena Penev, Kristina Vušković
Main reference Frédéric Maffray, Irena Penev, Kristina Vušković: "Coloring rings", CoRR, Vol. arXiv:1907.11905, 2019
URL https://arxiv.org/abs/1907.11905

A *ring* is a graph R whose vertex set can be partitioned into $k \geq 4$ nonempty sets $X_1, \ldots, X_k$ such that for all $i \in \{1, \ldots, k\}$ the set X_i can be ordered as $X_i = \{u_i^1, \ldots, u_i^{|X_i|}\}$ so that

$$X_i \subseteq N_R[u_i^{|X_i|}] \subseteq \cdots \subseteq N_R[u_i^1] = X_{i-1} \cup X_i \cup X_{i+1},$$

with subscripts taken modulo k. Under such circumstances, we say that the ring R is of *length* k. An *odd* (resp. *even*) ring is a ring of odd (resp. even) length.

Truemper configurations are prisms, pyramids, thetas, and wheels. Rings have played an important role in the study of a couple of classes defined by excluding certain Truemper configurations as induced subgraphs. A maximum clique and a maximum stable set of a ring can be computed in polynomial time, as can an optimal vertex-coloring of an even ring. However, odd rings present obstacles for coloring.

Our main result is that every ring R satisfies

$$\chi(R) = \max\{\chi(H) \mid H \text{ is a hyperhole in } R\}.$$

We present several corollaries of this result. One corollary is that the chromatic number of a ring can be computed in polynomial time.

3.13 Tutorial on Potential Maximal Cliques

Marcin Pilipczuk (University of Warsaw, PL)

In the tutorial, I presented the framework of finding maximum (weighted) independent set via minimal chordal completions and potential maximal cliques. The talk contained most of the details of the polynomial-time algorithm in the class of P_5-free graphs (Lokshtanov, Vatshelle, Villanger [2]). I also highlighted the main contribution of the seminal work of Bouchitté and Todinca [1] and difficulties in generalizing from P_5-free graphs to P_6-free graphs [3].

References

1 V. Bouchitté and I. Todinca. Treewidth and Minimum Fill-in: Grouping the minimal separators. *SIAM J. Comput.*, 31(1):212–232, 2001.
2 D. Lokshtanov, M. Vatshelle, and Y. Villanger. Independent set in P_5-free graphs in polynomial time. In C. Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 570–581. SIAM, 2014.
3 A. Grzesik, T. Klimošová, M. Pilipczuk, and M. Pilipczuk. Polynomial-time algorithm for maximum weight independent set on P_6-free graphs. In T. M. Chan, editor, *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 1257–1271. SIAM, 2019.

3.14 C_k-coloring of F-free graphs

Paweł Rzążewski (Warsaw University of Technology, PL)

Joint work of Maria Chudnovsky, Shenwei Huang, Paweł Rzążewski, Sophie Spirkl, Mingxian Zhong
Main reference Maria Chudnovsky, Shenwei Huang, Paweł Rzążewski, Sophie Spirkl, Mingxian Zhong: "Complexity of C_k-Coloring in Hereditary Classes of Graphs", in Proc. of the 27th Annual European Symposium on Algorithms (ESA 2019), LIPIcs, Vol. 144, pp. 31:1–31:15, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019.
URL http://dx.doi.org/10.4230/LIPIcs.ESA.2019.31

For a graph F, a graph G is *F-free* if it does not contain an induced subgraph isomorphic to F. For two graphs G and H, an *H-coloring* of G is a mapping $f : V(G) \to V(H)$ such that for every edge $uv \in E(G)$ it holds that $f(u)f(v) \in E(H)$. We are interested in the complexity of the problem H-Coloring, which asks for the existence of an H-coloring of an input graph G. In particular, we consider H-Coloring of F-free graphs, where F is a fixed graph and H is an odd cycle of length at least 5. This problem is closely related to the well known open problem of determining the complexity of 3-Coloring of P_t-free graphs.

We show that for every odd $k \geq 5$ the C_k-Coloring problem, even in the precoloring-extension variant, can be solved in polynomial time in P_9-free graphs. On the other hand, we prove that the extension version of C_k-Coloring is NP-complete for F-free graphs whenever some component of F is not a subgraph of a subdivided claw.

References

1 M.Chudnovsky, S. Huang, P. Rzążewski, S. Spirkl, M. Zhong, Complexity of C_k-coloring in hereditary classes of graphs, to appear in ESA 2019 Proc.

3.15 Polynomial Chi-binding functions and forbidden induced subgraphs: A survey

Ingo Schiermeyer (TU Bergakademie Freiberg, DE)

A graph G with clique number $\omega(G)$ and chromatic number $\chi(G)$ is *perfect* if $\chi(H) = \omega(H)$ for every induced subgraph H of G. A family $\mathcal{G}$ of graphs is called χ-bounded with binding function f if $\chi(G') \leq f(\omega(G'))$ holds whenever $G \in \mathcal{G}$ and G' is an induced subgraph of G. In this talk we will present a survey on polynomial χ-binding functions. Especially we will address perfect graphs, hereditary graphs satisfying the Vizing bound ($\chi \leq \omega + 1$), graphs having linear χ-binding functions and graphs having non-linear polynomial χ-binding functions. Thereby we also survey polynomial χ-binding functions for several graph classes defined in terms of forbidden induced subgraphs, among them $2K_2$-free graphs, P_k-free graphs, *claw*-free graphs, and *diamond*-free graphs.

3.16 Detecting an odd hole

Paul Seymour (Princeton University, US)

Joint work of Maria Chudnovsky, Alex Scott, Paul Seymour, Sophie Spirkl

A *hole* is a graph is an induced subgraph of length at least four, and an *antihole* is a hole in the complement. Odd holes are of particular interest, because of the strong perfect graph theorem, that says a graph is perfect if and only if it has no odd hole or odd antihole. A poly-time algorithm to test if a graph has an odd hole or odd antihole was found in 2006 [1], but detecting an odd hole, without stopping on discovery of an odd antihole, has remained open. We have now found a poly-time algorithm to test for odd holes [2]. Its running time is the same as the old algorithm, but in fact the details are much simpler.

References

1 M. Chudnovsky, G. Cornuéjols, X. Liu, P. Seymour and K. Vušković, "Recognizing Berge graphs", *Combinatorica*, 25 (2005), 143-186.
2 M. Chudnovsky, A. Scott. P. Seymour and S. Spirkl, "Detecting an odd hole", submitted for publication, `arXiv 1903.00208`

3.17 4-coloring P_6-free graphs

Sophie Spirkl (Rutgers University – Piscataway, US)

Joint work of Maria Chudnovsky, Mingxian Zhong, Sophie Spirkl

I talked about a recent polynomial-time algorithm for deciding if a given graph with no induced six-vertex path is four-colorable, and I discussed some of the methods used in the proof.

This is joint work with Maria Chudnovsky and Mingxian Zhong.

3.18 3-coloring with forbidden paths and cycles

Maya Jakobine Stein (University of Chile – Santiago de Chile, CL)

Joint work of Maya Jakobine Stein, Alberto Rojas

Graph coloring is hard, even if the number of colors is fixed. Therefore much effort has gone into determining the complexity of k-coloring special classes of graphs, in particular H-free graphs, where H is a fixed graph (and k is also fixed). It turns out that the problem remains NP-complete whenever H is not a linear forest, and for $k \geq 4$, the complexity of k-coloring P_t-free graphs has been determined for all values of t. There are polynomial time algorithms for 3-coloring P_t-free graphs for $t \leq 7$, but it is not known if such algorithms exist for $t = 8, 9, 10, \ldots$.. The algorithm given in [1] for P_7-free graphs was found by improving an earlier version which only worked for (P_7, C_3)-free graphs, so it seems natural to attack the problem by excluding one or more cycles in addition to the path P_t. We found a polynomial time algorithm for 3-coloring (P_9, C_5, C_3)-graphs. A variation of this algorithm works for all graphs having no induced P_{2t+1}, no induced odd cycle of length up to $2t-1$, and no induced C_8.

References

1 Bonomo, Flavia; Chudnovsky, Maria; Maceli, Peter; Schaudt, Oliver; Stein, Maya; Zhong, Mingxian *Three-Coloring and List Three-Coloring of Graphs Without Induced Paths on Seven Vertices.*Combinatorica (2018) 38: 779. https://doi.org/10.1007/s00493-017-3553-8

3.19 Layered wheels

Nicolas Trotignon (ENS – Lyon, FR)

Joint work of Ni Luh Dewi Sintiari, Nicolas Trotignon
Main reference Ni Luh Dewi Sintiari, Nicolas Trotignon: "(Theta, triangle)-free and (even hole, K_4)-free graphs. Part 1 : Layered wheels", CoRR, Vol. abs/1906.10998, 2019.
URL https://arxiv.org/abs/1906.10998

We present a construction called layered wheel. Layered wheels are graphs of arbitrarily large treewidth and girth. They might be an outcome for a possible theorem characterizing graphs with large treewidth in term of their induced subgraphs (while such a characterization is well understood in term of minors). They also provide examples of graphs of large treewidth and large rankwidth in well studied classes, such as (theta, triangle)-free graphs and even-hole-free graphs with no K_4 (where a hole is a chordless cycle of length at least 4, a theta is a graph made of three internally vertex disjoint paths of length at least 2 linking two vertices, and K_4 is the complete graph on 4 vertices).

3.20 Vertex colorings of interval hypergraphs

Zsolt Tuza (Alfréd Rényi Institute of Mathematics – Budapest, HU)

Joint work of Csilla Bujtás, Zsolt Tuza
Main reference Csilla Bujtás, Zsolt Tuza: "Color-bounded hypergraphs, VI: Structural and functional jumps in complexity", Discrete Mathematics, Vol. 313(19), pp. 1965–1977, 2013.
URL https://doi.org/10.1016/j.disc.2012.09.020

The classical notion of proper coloring requires a color assignment to the vertices in such a way that no hyperedge is monochromatic (Erdős & Hajnal, mid-1960's). Equivalently this means at least two colors in each hyperedge. In C-coloring the restriction is put from the other side, namely that every hyperedge e is allowed to contain at most $|e|-1$ colors (Berge / Sterboul, early 1970's). In the more complex model of mixed hypergraps both types of hyperedges may occur (Voloshin, early 1990's). A generalization of this structure class is obtained by putting lower and/or upper bounds on the largest cardinality of monochromatic subsets of – and/or on the number of colors occurring in – each hyperedge (Bujtás & Tuza, mid-2000's).

Despite that lots of results are known, some simple questions are still open, even on interval hypergraphs. (An interval hypergraph is a collection of hyperedges $e_1, \ldots, e_m$ whose underlying vertex set admits an ordering such that each hyperedge e_i consists of consecutive vertices without gap in that order.) As an example of problems unsolved for over a decade, assume that for each e_i there is a color which appears on at least a given number a_i of vertices inside e_i. The task is to determine the largest possible number of colors. Is this optimization problem polynomial-time solvable or NP-hard?

In the talk we mention open questions concerning the chromatic polynomial, too.

4 Open problems

4.1 Parameterized complexity of the coloring problems for H-free graphs

Petr A. Golovach (University of Bergen, NO)

Only very few parameterized results for Coloring on H-free graphs are known. We refer to [2] for the detailed survey of the known results and open problems. Here we underline two problems that we believe to be the most interesting.

It is a very long standing open problem whether the 3-Coloring problem for P_ℓ-free graphs admits a polynomial algorithm for every positive ℓ or it becomes NP-complete for some $\ell \geq 8$. Currently, it is known that the problem can be solved in polynomial time for $\ell \leq 7$ [1]. This leads to the following question.

- Is 3-Coloring W[1]-hard on P_ℓ-free graphs when parameterized by ℓ?

The next problem was first stated by Hoàng et al. [3]. They proved that that k-Coloring can be solved in polynomial time on P_5-free graphs for every positive integer k, that is the problem is in XP when parameterized by k, but left open the question whether there is a matching lower bound or their result may be improved.

- Is k-Coloring FPT on P_5-free graphs when parameterized by k?

The question whether k-Coloring parameterized by k is FPT is also open and interesting for $2P_2$-free graphs that compose a subclass of P_5-free graphs.

References

1 F. Bonomo, M. Chudnovsky, P. Maceli, O. Schaudt, M. Stein, and M. Zhong, *Three-coloring and list three-coloring of graphs without induced paths on seven vertices*, Combinatorica, 38 (2018), pp. 779–801.
2 P. A. Golovach, M. Johnson, D. Paulusma, and J. Song, *A survey on the computational complexity of coloring graphs with forbidden subgraphs*, Journal of Graph Theory, 84 (2017), pp. 331–363.
3 C. T. Hoàng, M. Kaminski, V. V. Lozin, J. Sawada, and X. Shu, *Deciding* k*-colorability of* P$_5$*-free graphs in polynomial time*, Algorithmica, 57 (2010), pp. 74–81.

4.2 The Dilworth number of a graphs

Chinh T. Hoàng (Wilfrid Laurier University – Waterloo, CA)

Given a graph G, a vertex x dominates a vertex y if every neighbor of y, different from x, is a neighbor of x. Vertex x is comparable to vertex y, if x dominates y, or y dominates x. The domination relation is a partial order. The Dilworth number of a graph G is the largest number of pairwise incomparable vertices in G.

Problem 1. Is it true that there is a polynomial time algorithm to optimally color all graphs with bounded Dilworth number?

Problem 2. Is it true that if a graph G has bounded Dilworth number then it has bounded clique width? After I posed this problem, it was pointed out to me that the answer is NO. The authors Korpelainen, Lozin, and Mayhill constructed a graph with Dilworth number two and with arbitrarily high clique width.
(https://link.springer.com/content/pdf/10.1007/s00373-013-1290-3.pdf)

4.3 Colouring Graphs of Bounded Diameter

Daniel Paulusma (Durham University, GB)

It is known that k-Colouring is NP-complete for graphs of diameter at most d for all pairs (k, d) with $k \geq 3$ and $d \geq 2$ except when $(k, d) = (3, 2)$: determining the computational complexity of 3-Colouring for graphs of diameter 2 is a long-standing open problem. The following related problems are also open:

Determine the computational complexity of 3-Colouring and Colouring (where k is part of the input) for triangle-free graphs of diameter 2.

It can be observed that for all integers $d, k, r \geq 1$, the k-Colouring problem is constant-time solvable for $K_{1,r}$-free graphs of diameter d and that Colouring is NP-complete for $K_{1,4}$-free graphs. However, the following problem is open:

Determine the computational complexity of Colouring restricted to $K_{1,3}$-free graphs of diameter d for every $d \geq 2$.

The above observations and open problems can all be found in [1].

References

1 Barnaby Martin, Daniël Paulusma and Siani Smith, Colouring H-free graphs of bounded diameter, Proceedings of the 44th International Symposium on Mathematical Foundations of Computer Science (MFCS 2019), Leibniz International Proceedings in Informatics, to appear.

4.4 Odd cycle transversal in P_5-free graphs

Paweł Rzążewski (Warsaw University of Technology, PL)

Joint work of Konrad K. Dabrowski, Carl Feghali, Matthew Johnson, Giacomo Paesani, Daniël Paulusma, Paweł Rzążewski
Main reference Konrad K. Dabrowski, Carl Feghali, Matthew Johnson, Giacomo Paesani, Daniël Paulusma, Paweł Rzążewski: "On Cycle Transversals and Their Connected Variants in the Absence of a Small Linear Forest", CoRR, Vol. abs/1908.00491, 2019.
URL http://arxiv.org/abs/1908.00491

It is known that the Odd Cycle Transversal is polynomial-time solvable in P_4-free graphs, but is NP-complete in P_6-free graphs [1]. What is the complexity of the problem in P_5-free graphs?

References

1 K. Dabrowski, C. Feghali, M. Johnson, G. Paesani, D. Paulusma, P. Rzążewski, *On Cycle Transversals and Their Connected Variants in the Absence of a Small Linear Forest*, manuscript

4.5 Computing disjoint paths

Nicolas Trotignon (ENS – Lyon, FR)

Consider the following problem:

Input: A graph G and a, b, c, d four vertices of G.

Question: Does there exist two paths, vertex disjoint, with no edges between them and such that their ends are all in $\{a, b, c, d\}$?

The complexity of this problem is not known.

Remarks:

- The problem is trivial when $\{a, b, c, d\}$ is not a stable set of size four. In a solution, each of a, b, c and d must be an end of exactly one of the paths.
- Deciding whether there exit a path from a to b and a path from c to d, vertex-disjoint and with edges between them is NP-complete as shown by Bienstock. So, to solve the problem, it is hopeless to try the three possible ways the paths may exist separately (a-b + c-d; a-c + b-d; a-d + b-c).

The problem is motivated by the detection of induced minors, see the references.

References

1 D. Bienstock. "On the complexity of testing for odd holes and induced odd paths", Discrete Math. 90 (1991) 85–92. (Corrigendum, Discrete Mathematics 102 (1992) 109.)
2 J. Fiala, M. Kaminski and D. Paulusma. "Induced containment relations in claw-free and general graphs", Journal of Discrete Algorithms 17 (2012) 74–85.

4.6 Precoloring extension of graphs

Zsolt Tuza (Alfréd Rényi Institute of Mathematics – Budapest, HU)

The decision problem t-PrExt is the following subproblem of Precoloring Extension.

Input: Graph $G = (V, E)$, nonnegative integer k, proper coloring of an induced subgraph $H \subset G$, such that each color occurs at most t times in H.
Question: Can the coloring of H be extended to a proper k-coloring of the entire G?

Clearly, 0-PrExt means k-colorability, hence it is linear-time solvable on bipartite graphs. However, already 1-PrExt is NP-complete on bipartite graphs [2]. Further, on interval graphs 1-PrExt is solvable in polynomial time, but 2-PrExt is NP-complete [1].

PROBLEM 1. For $t > 1$, find graph classes in which t-PrExt is polynomial-time solvable and $(t + 1)$-PrExt is NP-complete.

PROBLEM 2. On interval graphs, design a linear-time algorithm for 1-PrExt, or prove that every 1-PrExt algorithm is superlinear in $|V| + |E|$.

References

1 M. Biró, M. Hujter, Zs. Tuza, "Precoloring extension. I. Interval graphs," Discrete Math. 100:267–279, 1992.
2 M. Hujter, Zs. Tuza, "Precoloring extension. II. Graph classes related to bipartite graphs," Acta Math. Univ. Comenianae, 62(1):1–11, 1993.

Participants

- Isolde Adler
University of Leeds, GB
- Marthe Bonamy
University of Bordeaux, FR
- Nicolas Bousquet
Grenoble INP, FR
- Christoph Brause
TU Bergakademie Freiberg, DE
- Kathie Cameron
Wilfrid Laurier University – Waterloo, CA
- Maria Chudnovsky
Princeton University, US
- Konrad Dabrowski
Durham University, GB
- Cemil Dibek
Princeton University, US
- Francois Dross
University of Montpellier & CNRS, FR
- Esther Galby
University of Fribourg, CH
- Petr A. Golovach
University of Bergen, NO
- Chinh T. Hoàng
Wilfrid Laurier University – Waterloo, CA
- Shenwei Huang
Nankai University – Tianjin, CN
- Bart Jansen
TU Eindhoven, NL
- Matthew Johnson
Durham University, GB
- Mamadou Moustapha Kanté
Université Clermont Auvergne – Aubiere, FR
- Tereza Klimosova
Charles University – Prague, CZ
- Stefan Kratsch
HU Berlin, DE
- O-joung Kwon
Incheon National University, KR
- Bernard Lidicky
Iowa State University, US
- Anita Liebenau
UNSW Sydney, AU
- Paloma Lima
University of Bergen, NO
- Daniel Lokshtanov
University of California – Santa Barbara, US
- Peter Maceli
Adelphi University – Garden City, US
- Tomáš Masařík
Charles University – Prague, CZ
- Sang-il Oum
IBS – Daejeon, KR
- Daniel Paulusma
Durham University, GB
- Irena Penev
Charles University – Prague, CZ
- Marcin Pilipczuk
University of Warsaw, PL
- Michal Pilipczuk
University of Warsaw, PL
- Bernard Ries
University of Fribourg, CH
- Paweł Rzążewski
Warsaw University of Technology, PL
- Oliver Schaudt
RWTH Aachen, DE
- Ingo Schiermeyer
TU Bergakademie Freiberg, DE
- Pascal Schweitzer
TU Kaiserslautern, DE
- Paul Seymour
Princeton University, US
- Sophie Spirkl
Rutgers University – Piscataway, US
- Juraj Stacho
Google Switzerland – Zürich, CH
- Maya Jakobine Stein
University of Chile – Santiago de Chile, CL
- Stéphan Thomassé
ENS – Lyon, FR
- Nicolas Trotignon
ENS – Lyon, FR
- Zsolt Tuza
Alfréd Rényi Institute of Mathematics – Budapest, HU
- Erik Jan van Leeuwen
Utrecht University, NL
- Gerhard J. Woeginger
RWTH Aachen, DE
- Victor Zamaraev
Durham University, GB

Report from Dagstuhl Seminar 19272

Real VR – Importing the Real World into Immersive VR and Optimizing the Perceptual Experience of Head-Mounted Displays

Edited by

Marcus A. Magnor[1] and Alexander Sorkine-Hornung[2]

1 **TU Braunschweig, DE,** magnor@cg.cs.tu-bs.de

2 **Facebook Zürich, CH,** alexsh@fb.com

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 19272 "Real VR – Importing the Real World into Immersive VR and Optimizing the Perceptual Experience of Head-Mounted Displays".

Motivated by the advent of mass-market VR headsets, this Dagstuhl Seminar addresses the scientific and engineering challenges that need to be overcome in order to experience omni-directional video recordings of the real world with the sense of stereoscopic, full-parallax immersion as can be provided by today's head-mounted displays.

Since the times of the Lumière brothers, the way we watch movies hasn't fundamentally changed: Whether in movie theaters, on mobile devices, or on TV at home, we still experience movies as outside observers, watching the action through a "peephole" whose size is defined by the angular extent of the screen. As soon as we look away from the screen or turn around, we are immediately reminded that we are only "voyeurs". With modern full-field-of-view, head-mounted and tracked VR displays, this outside-observer paradigm of visual entertainment is quickly giving way to a fully immersive experience. Now, the action fully encompasses the viewer, drawing us in much more than was possible before.

For the time being, however, current endeavors towards immersive visual entertainment are based almost entirely on 3D graphics-generated content, limiting application scenarios to purely digital, virtual worlds only. The reason is that in order to provide for stereo vision and ego-motion parallax, which are both essential for genuine visual immersion perception, the scene must be rendered in real-time from arbitrary vantage points. While this can be easily accomplished with 3D graphics via standard GPU rendering, it is not at all straight-forward to do the same from conventional video footage acquired of real-world events.

Another challenge is that consumer-grade VR headsets feature spatial resolutions that are still considerably below foveal acuity, yielding a pixelated, subpar immersive viewing experience. At the same time, the visual perception characteristics of our fovea are decidedly different from our peripheral vision (as regards spatial and temporal resolution, color, contrast, clutter disambiguation etc.). So far, computer graphics research has focused almost entirely on foveal perception, even though our peripheral vision accounts for 99% of our field of view. To optimize perceived visual quality of head-mounted immersive displays, and to make optimal use of available computational resources, advanced VR rendering algorithms need to simultaneously account for our foveal and peripheral vision characteristics.

The aim of the seminar was to collectively fathom what needs to be done to facilitate truly immersive viewing of real-world recordings and how to enhance the immersive viewing experience by taking perceptual aspects into account. The topic touches on research aspects from various fields, ranging from digital imaging, video processing, and computer vision to computer graphics, virtual reality, and visual perception. The seminar brought together scientists, engineers and practitioners from industry and academia to form a lasting, interdisciplinary research community who set out to jointly address the challenges of Real VR.

Real VR – Importing the Real World into Immersive VR and Optimizing the Perceptual Experience of Head-Mounted Displays, *Dagstuhl Reports*, Vol. 9, Issue 6, pp. 143–156

Editors: Marcus A. Magnor and Alexander Sorkine-Hornung

DAGSTUHL REPORTS Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Seminar June 30 – July 3, 2019 – http://www.dagstuhl.de/19272
2012 ACM Subject Classification Computing methodologies → Computer graphicsComputing methodologies → Computer vision
Keywords and phrases immersive digital reality, perception in vr, real-world virtual reality
Digital Object Identifier 10.4230/DagRep.9.6.143
Edited in cooperation with Susana Castillo

1 Executive Summary

Marcus A. Magnor (TU Braunschweig, DE)
Alexander Sorkine-Hornung (Facebook Zürich, CH)

The Dagstuhl seminar brought together 27 researchers and practitioners from academia and industry to discuss the state-of-the-art, current challenges, as well as promising future research directions in Real VR. Real VR, as defined by the seminar participants, pursues two overarching goals: facilitating the import of real-world scenes into head-mounted displays (HMDs), and attaining perceptual realism in HMDs. The vision of Real VR is enabling to experience movies, concerts, even live sports events in HMDs with the sense of immersion of really "being-there", unattainable by today's technologies.

In the welcome and overview session, the participants collectively decided on the seminar program for the following days. In total, the seminar program included the overview session, three research presentation sessions, two breakout sessions including a demo track, two sessions for one-on-one discussions and individual exchange, one session for writing up the results, plus the summary and closing session.

To kick off the seminar, Alexander Sorkine-Hornung from Oculus VR presented the latest developments from an industrial perspective. He gave insights from the development of the just-released Oculus Quest and Oculus Rift S HMDs. In the research presentation sessions, 21 participants gave talks on their work. Participants also met in smaller groups in the breakout sessions to discuss the specific challenges of these fields in more detail. In due course, it became apparent that Real VR concerns research challenges in a number of different fields:

- Capture
- Reconstruction & modeling
- Rendering & perception
- Display technologies
- Interaction & virtual avatars
- Production & applications

Some exemplary results of the seminar on these topics were:

The persistent lack of consumer-market, i.e. affordable, mid- to high-resolution 360-degree video cameras to **capture** dynamic real-world scenes omnidirectionally still hamper research and development in Real VR. So far, research groups largely build their own custom-designed omnidirectional video cameras. Prominent examples include the omnidirectional camera designs by the group of Philippe Bekaert from Hasselt University, Belgium, and the top-of-the-line Manifold camera presented by Brian Cabral from Facebook. Besides novel devices, also simpler recording methods are sought, e.g. by Tobias Bertel and Christian Richardt at Bath, in order to capture real-world content more casually.

On **scene reconstruction and representation**, the jury is still out whether omnidirectional video should be considered to represent sparse light field data with dense depth/disparity as side information, or whether panoramic footage should (and could) be processed to provide full 3D geometry representations of the scene. As pointed out by Atanas Gotchev from TU Tampere, Marco Volino from the University of Surrey, and Christian Richardt from the University of Bath, both forms of representation have their respective advantages and drawbacks, e.g. when aiming to augment the real scene with additional virtual content.. Memory requirements and real-time streaming bandwidth requirements are challenging in either case.

The form of scene representation also determines which **rendering** approaches are viable. For 3D rendering, Dieter Schmalstieg from Graz presented his Shading Atlas Streaming approach to efficiently divide shading and rendering computation between server and client. To make use of visual **perception** characteristics in wide field-of-view HMDs, on the other hand, foveated rendering approaches, e.g. based on hardware ray tracing and accelerated machine learning, as presented by Anjul Patney from NVidia, have great potential. As shown by Qi Sun from Adobe, perceptual methods like saccade-aware rendering can also be used to enable walking through huge virtual worlds while actually not leaving the confines of one's living room. To render from dense depth-annotated 360-deg video, in contrast, advanced image-based warping methods and hole-filling approaches are needed, as was convincingly outlined by Tobias Bertel from the University of Bath.

Gordon Wetzstein from Stanford University presented how future HMDs will become even more realistic by overcoming current limitations of near-eye **displays**, in particular the vergence-accommodation conflict. Along similar lines, Hansung Kim from the University of Surrey showed how spatial audio enhances perceived VR realism even more.

Social **interaction** in the virtual world requires having digital doubles available. The elaborate steps needed to create convincing human avatars from real-world people were outlined by Feng Xu from Tsinghua University, Darren Cosker from the University of Bath, Christian Theobalt from MPII, and Peter Eisert from TU Berlin, covering the full range of human face, hand and body capture, reconstruction, and modeling. To interact with objects in virtual space, on the other hand, Erroll Wood from Microsoft Cambridge described how hand motion and gestures can be reliably tracked and identified in real-time by the upcoming HoloLens 2 device. Also based on real-time tracking, Li-Yi Wei from Adobe presented a system that enables presenters to augment their live presentation by interacting with the shown content in real-time using mere hand gestures and body postures.

Regarding **content production and applications**, Christian Lipski from Apple presented the ARKit software framework developed for creating captivating augmented reality experiences. James Tompkin from Brown University presented work on multi-view camera editing of Real VR content during post-production. Johanna Pirker from TU Graz showed how virtual reality can be paired with human-computer interaction to enhance learning experiences in the physics classroom. Production aspects and cinematic VR experiences were also considered prominent drivers of contemporary Real VR research by other presenters, e.g. Marco Volino, Darren Cosker, Philippe Bekaert, Peter Eisert and Brian Cabral.

Practically experiencing the new, tetherless Oculus Quest brought along by Alexander Sorkine-Hornung in the **demonstration track** made impressively clear how free, unrestricted user motion extends the usability and acceptance of VR tremendously, made possible by the pass-through view feature of this HMD.

Finally, in the coming months, a number of seminar participants will compile an edited book volume on the state-of-the-art in Real VR that Springer has already agreed to publish as part of their well-known Lecture Notes on Computer Science (LNCS) Survey Series.

2 Table of Contents

3 Overview of Talks

3.1 Bridging the gap between VR and TV

Philippe Bekaert (Hasselt University – Diepenbeek, BE)

In this talk, I present a brief overview of VR/AR related R&D work at the visual computing group of the Expertise center for Digital Media of Hasselt University in Diepenbeek, Belgium, and our young spin-off AZilPix. Topics covered include 360 degrees video camera rigs and stitching, free viewpoint video, and a range of applications we have been addressing: immersive collaboration, free viewpoint video for sports broadcasting, video capture for multi-scopic ("holographic") video displays, researching the "grammar" of the new medium which immersive video implies (with the art performance company CREW, www.crewonline.org), investigating the use of 360 video as "the next best thing to being there" (e.g. at major rock festivals), an immersive video environment for medical surgery training (with another young company dlivemed, www.dlivemed.com), and the use of immersive video as a conventional TV broadcasting production tool (the topic of our AZilPix spin-off, www.azilpix.com).

AZilPix commercialises a multi-camera video capture and processing system offering virtual pan-tilt-zoom functionality in ultra high resolution (48 megapixels) cameras. It is rooted in our expertise with 360 degree video systems, and offers the ability to mix 360 video and regular broadcast TV video in one and the same system. It offers entirely stitching artifact free TV broadcast quality 360 degrees video in a very simple and convenient way: by means of an extreme fish eye lens on our ultra-high resolution camera.

3.2 Casual Real VR

Tobias Bertel (University of Bath, GB)

Joint work of Tobias Bertel, Neill D.F. Campbell, Christian Richardt

Creating Real VR experiences is a very challenging and multi-disciplinary task. Real-world capturing procedures as well as realistic scene representations are hard to obtain for general environments and are key for high-quality novel-view synthesis. My talk begins by motivating image-based rendering (IBR) techniques. I focus on the 360º representation and talk about its issues. My work, MegaParallax, which provides 360º environments with motion parallax, does a step forward to provide more immersive Real VR experiences in ego-centric environments. Concepts for the current state-of-the-art in traditional and learned novel-view synthesis, namely layered scene geometry, are motivated and illustrated. My talk concludes by pointing to encountered issues during my research dealing with 360º environments and sharing a set of ideas for the shiny future, which are mainly orbiting around reliable 3D reconstruction for ego-centric camera paths, which are crucial for casual Real VR and AR applications, and deep learning.

3.3 Full-360 Inside-out Volumetric Video Capture. Six Degrees of Freedom Video.

Brian Cabral (Facebook – Menlo Park, US)

Using a single camera array to capture depth + RGB color remains a daunting challenge for a variety of reasons ranging from the physical creation of a reliable camera rig, to managing the large data sets, and then to the reconstruction of a final 360 image. As part of our research and development we built such a system based on cinematic quality cameras and sensors. We present the HW and SW architecture, key algorithms, and technical hurdles we overcame. The talk will conclude with a presentation of results and remaining challenges.

3.4 Virtual Characters and Interaction in Immersive Worlds

Darren Cosker (University of Bath, GB)

We are now beginning to see high quality real time characters suitable for immersive experiences. However, these typically one-off exercises require a great deal of effort, cost and expertise to create. In this talk I will outline some recent exciting examples of work in this area – including recent CAMERA work which contributed to commercial VR and video game projects with the BBC and Aardman. I will draw out some of our and our commercial partners experiences in this area – applicable to many different aspects of virtual character creation and animation – which I hope will lead to interesting future challenges.

3.5 A Perceptual Graphics Approach to Non-Verbal Communication

Douglas Cunningham (BTU Cottbus-Senftenberg, DE)

By integrating the knowledge and methods from the fields of Psychology and Informatics, it is possible to address research questions that were otherwise intractable. In particular, my lab combines low-level psychophysics with modern computer graphics to allow an ecological psychology approach to modeling and synthesizing behaviorally relevant, spatio-temporal information. In this talk, I will provide an overview of recent work on the perception and (artificial) synthesis of facial expressions as well as work on the relationship between low-level image statistics and perception.

3.6 Hybrid Human Modeling

Peter Eisert (Fraunhofer-Institut – Berlin, DE)

Photorealistic modeling and rendering of humans is extremely important for VR environments, as the human body and face are highly complex and exhibit large shape variability but also especially as humans are extremely sensitive to looking at humans. Furthermore, in VR environments, interactivity plays an important role. While purely computer graphics modeling can achieve highly realistic human models, achieving real photorealism with these models is extremely computationally expensive. In our research we investigate hybrid representations for human bodies and faces, combining classical computer graphics models with image- and video-based as well as example-based approaches in order to combine interactivity with photorealism. In this talk, we will cover recent advances at HHI in 4D human body reconstruction and modeling as well as hybrid face modeling.

3.7 3D Visual Scene Sensing, Light Field Processing and Immersive Visualization

Atanas Gotchev (Tampere University of Technology, FI)

This talk summarizes three topics of our research placed along the chain of 3D visual scene sensing, light processing and immersive visualization.

First, we discuss the problem of denoising 3D scene range measurements acquired by Time-of-flight (ToF) range sensors and composed in the form of 2D image-like depth maps. We address the specific case of ToF low-sensing environment, which is set by low-light sensing conditions, low-power hardware requirements, and low-reflectivity scenes. We present an elaborated analysis of noise properties of ToF data sensed in low-sensing conditions and a related non-local denoising approach working in complex domain.

Second, we discuss the problem of light field reconstruction through sparse modelling in shearlet domain. We argue that the generated densely sampled light field of a given 3D scene is suitable for all applications which require light field reconstruction.

Third, we discuss our practical experience in developing a wide super-multiview head-up display, which augment the vehicle windshield with images flying few meters in front of the driver and maintaining smooth parallax.

3.8 Spatial Audio Reproduction System for VR Using 360° Cameras

Hansung Kim (University of Surrey – Guildford, GB)

Recent progresses in Virtual Reality (VR) and Augmented Reality (AR) allow us to experience various VR/AR applications in our daily life. In order to maximise the immersion of user in VR/AR environments, a plausible spatial audio reproduction synchronised with visual information is essential. In this talk, we introduce a simple and efficient system to estimate room acoustic for plausible reproduction of spatial audio using 360° cameras for VR/AR applications. A pair of 360° images is used for room geometry and acoustic property estimation. A simplified 3D geometric model of the scene is estimated by depth estimation from captured images and semantic labelling using a convolutional neural network (CNN). The real environment acoustics are characterised by frequency-dependent acoustic predictions of the scene. Spatially synchronised audio is reproduced based on the estimated geometric and acoustic properties in the scene. The reconstructed scenes are rendered with synthesised spatial audio as VR/AR content.

3.9 Introducing Apple's ARKit 3: Real-Time Computer Vision on Mobile Devices

Christian Lipski (Apple Computer Inc. – Cupertino, US)

ARKit is a cutting-edge software framework that makes it easy for developers to create captivating augmented reality experiences for iPhone and iPad. The ARKit team performs research in the field of computer vision/machine learning, on various topics such as visual-inertial slam, face tracking, scene reconstruction and scene understanding. I will present some of the latest advancements and show which tracking and scene reconstruction algorithms run today on our mobile devices. Furthermore, I will address the challenges of ensuring consistent, high quality results of these algorithms.

3.10 From Reality to Immersive VR. What's missing in VR?

Marcus A. Magnor (TU Braunschweig, DE)

Current endeavors towards immersive visual entertainment are still almost entirely based on 3D graphics-generated content, limiting application scenarios to virtual worlds only. The reason is that in order to provide for stereo vision and ego-motion parallax, two essential ingredients for genuine visual immersion perception, the scene must be rendered in real-time from arbitrary vantage points. While this can be easily accomplished with 3D graphics via standard GPU rendering, it is not at all straight-forward to do the same from conventional video footage acquired of real-world events. In my talk I will outline avenues of research toward enabling the immersive experience of real-world recordings and how to enhance the immersive viewing experience by taking perceptual issues into account.

I will report on the latest results of our project "Immersive Digital Reality" which is funded by the German Science Foundation (DFG MA2555/15-1).

3.11 Towards Deep Real-time Rendering for Mixed Reality

Anjul Patney (Facebook – Redmond, US)

Advances in real-time graphics have enabled some of today's most immersive visual experiences. However, recent disruptions have transformed the landscape of technology in this area. On one hand, increasing demands from mixed-reality and other displays have forced us to rethink how we render real-time graphics for high resolutions, framerates, and fields of view. On the other hand, hardware ray tracing and accelerated machine learning have provided two promising opportunities to address these challenges. The result is an exciting opportunity to shape the next generation of interactive computer graphics. In my talk I will present how the two directions fit into my vision of the future of mixed reality graphics.

3.12 HCI meets VR: Learning through VR

Johanna Pirker (TU Graz, AT)

3.13 6-DoF VR Video: Towards Immersive 360° VR Video with Motion Parallax

Christian Richardt (University of Bath, GB)

Main reference Tobias Bertel, Neill D. F. Campbell, Christian Richardt: "MegaParallax: Casual 360° Panoramas with Motion Parallax", IEEE Trans. Vis. Comput. Graph., Vol. 25(5), pp. 1828–1835, 2019.
URL http://dx.doi.org/10.1109/TVCG.2019.2898799

Virtual reality is a new kind of medium that requires new ways to author content. My goal is therefore to create a new form of immersive 360-degree VR video that overcomes the limitations of existing 360-degree VR video. This new form of VR content – 6-DoF VR video – will achieve unparalleled realism and immersion by providing freedom of head motion and motion parallax, which is a vital depth cue for the human visual system and entirely missing from existing 360-degree VR video. Specifically, my aim is to accurately and comprehensively capture real-world environments, including visual dynamics such as people and moving animals or plants, and to reproduce the captured environments and their dynamics in VR with photographic realism, correct motion parallax and overall depth perception. 6-DoF VR video is a significant virtual reality capability that will be a significant step forward for overall immersion, realism and quality of experience.

3.14 Shading Atlas Streaming

Dieter Schmalstieg (TU Graz, AT)

Main reference Joerg H. Mueller, Philip Voglreiter, Mark Dokter, Thomas Neff, Mina Makar, Markus Steinberger, Dieter Schmalstieg: "Shading atlas streaming", ACM Trans. Graph., Vol. 37(6), pp. 199:1–199:16, 2018.
URL http://dx.doi.org/10.1145/3272127.3275087

Streaming high quality rendering for virtual reality applications requires minimizing perceived latency. We introduce Shading Atlas Streaming (SAS), a novel object-space rendering framework suitable for streaming virtual reality content. SAS decouples server-side shading from client-side rendering, allowing the client to perform framerate upsampling and latency compensation autonomously for short periods of time. The shading information created by the server in object space is temporally coherent and can be efficiently compressed using standard MPEG encoding. Our results show that SAS compares favorably to previous methods for remote image-based rendering in terms of image quality and network bandwidth efficiency. SAS allows highly efficient parallel allocation in a virtualized-texture-like memory hierarchy, solving a common efficiency problem of object-space shading. With SAS, untethered virtual reality headsets can benefit from high quality rendering without paying in increased latency.

3.15 Human Learning: Understanding and Computing the Eyes and Brain in VR

Qi Sun (Adobe Inc. – San José, US)

Immersive content authoring and consumption are critical components for the next generation VR platforms. It includes real-time rendering/interface software, and display hardware. In this talk, I will present my previous research on understanding and (more importantly) leveraging human perceptual factors in the eye and brain for fundamental geometry/imaging/rendering algorithms and practical applications, such as VR walk-through and computationally foveated displays.

3.16 Capturing the real world for interaction in VR and AR

Christian Theobalt (MPI für Informatik – Saarbrücken, DE)

In this talk, I will briefly review some of our recent work on high quality human face, body and hand capture, as well as inverse rendering with a single color or depth camera.

3.17 Multi-view Camera Editing: A Few Notes

James Tompkin (Brown University – Providence, US)

Capturing the world realistically for VR reproduction is often only the start of the creative process: editing the data comes next. In my talk, I discussed some of the problems with editing multi-camera footage and presented some of my recent work in this area. These works attempt to build a set of tools, similar to how Adobe Photoshop can edit 2D data, but for multi-camera footage. For instance, I discussed how we can filter multi-view videos in a spatio-temporally consistent way. Then, I discussed how we can perform region selection in multi-view image data via view-consistent superpixels and scribble selection. I also discussed future display devices for displaying VR content in social settings using massive tiled displays, e.g., for virtual concerts.

3.18 Light Fields for Immersive Virtual Experiences

Marco Volino (University of Surrey, GB)

There is increasing interest in creating photo-realistic cinematic experiences for virtual and augmented reality (AR/VR). The use of games engines and computer generated imagery allows inter-activity, such as head movement, but do not achieve cinematic photo-realism. Conversely, 360 and stereo 360 video achieve photo-realism for a fixed location but do not allow head movement or realistic parallax. Light field capture offers a potential solution for capturing a scene with full photo-realism whilst allowing changes in viewpoint with correct parallax. This talk will discuss recent work for the creation of light field video assets for use in cinematic VR experiences.

3.19 Interactive Body-driven Graphics for Augmented Video Performance

Li-Yi Wei (Adobe Inc. – San José, US)

Joint work of Nazmus Saquib, Rubaiat Habib Kazi, Li-Yi Wei, Wilmot Li
Main reference Nazmus Saquib, Rubaiat Habib Kazi, Li-Yi Wei, Wilmot Li: "Interactive Body-Driven Graphics for Augmented Video Performance", in Proc. of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019, p. 622, ACM, 2019.
URL https://doi.org/10.1145/3290605.3300852

We naturally use hand gestures and body postures when talking to one another, but often resort to pre-created slides or videos when presenting with synthetic or virtual effects. Is there a way to combine the spontaneity of natural gestures with the power of synthetic graphics to enrich our presentation and communication? We investigate a prototype system that augments live presentations with interactive graphics to create a rich and expressive storytelling environment. Using our system, the presenter interacts with the graphical elements in real-time with gestures and postures, thus leveraging our innate, everyday skills to enhance our communication capabilities with the audience.

3.20 Computational Near-eye Displays & Digital Eyeglasses with Focus Cues

Gordon Wetzstein (Stanford University, US)

Immersive visual and experiential computing systems, i.e. virtual and augmented reality (VR/AR), are entering the consumer market and have the potential to profoundly impact our society. Applications of these systems range from communication, entertainment, education, collaborative work, simulation and training to telesurgery, phobia treatment, and basic vision research. In every immersive experience, the primary interface between the user and the digital world is the near-eye display. Thus, developing near-eye display systems that provide a high-quality user experience is of the utmost importance. Many characteristics of near-eye displays that define the quality of an experience, such as resolution, refresh rate, contrast, and field of view, have been significantly improved over the last years. However, a significant source of visual discomfort prevails: the vergence-accommodation conflict (VAC). Further, natural focus cues are not supported by any existing near-eye display. In this talk, we discuss frontiers of engineering next-generation opto-computational near-eye display systems to increase visual comfort and provide realistic and effective visual experiences.

3.21 Hand Tracking for HoloLens 2

Erroll Wood (Microsoft Research – Cambridge, GB)

Traditional computing devices require us to adapt ourselves to digital input and output mechanisms, mediating our interactions through flat screens, button presses and touch events. Mixed reality devices like HoloLens open up a new computing paradigm, where digital content integrates seamlessly with the physical world, and our devices understand the environments we inhabit and the people that inhabit them. I will describe how HoloLens 2 tracks the user's hands, allowing the user to interact with virtual objects by simply reaching out and touching them.

3.22 Light Weight 3D Dynamic Reconstruction

Feng Xu (Tsinghua University Beijing, CN)

3D dynamic reconstruction is an important topic in graphics and computer vision. This talk majorly focuses on the reconstruction of human face, body and hand, which have many applications in HCI, VR and AR. The techniques use one or two consumer level sensors to reconstruct the motions of the target in real time, which we believe may have a great potential to be used in the future mobile devices and applications for end users.

Participants

- Philippe Bekaert
Hasselt University –
Diepenbeek, BE
- Tobias Bertel
University of Bath, GB
- Brian Cabral
Facebook – Menlo Park, US
- Susana Castillo Alejandre
TU Braunschweig, DE
- Darren Cosker
University of Bath, GB
- Douglas Cunningham
BTU Cottbus-Senftenberg, DE
- Peter Eisert
Fraunhofer-Institut – Berlin, DE
- Atanas Gotechev
Tampere University of
Technology, FI
- Adrian Hilton
University of Surrey –
Guildford, GB
- Moritz Kappel
TU Braunschweig, DE
- Hansung Kim
University of Surrey –
Guildford, GB
- Christian Lipski
Apple Computer Inc. –
Cupertino, US
- Marcus A. Magnor
TU Braunschweig, DE
- Anjul Patney
Facebook – Redmond, US
- Johanna Pirker
TU Graz, AT
- Christian Richardt
University of Bath, GB
- Dieter Schmalstieg
TU Graz, AT
- Alexander Sorkine-Hornung
Facebook Zürich, CH
- Frank Steinicke
Universität Hamburg, DE
- Qi Sun
Adobe Inc. – San José, US
- Christian Theobalt
MPI für Informatik –
Saarbrücken, DE
- James Tompkin
Brown University –
Providence, US
- Marco Volino
University of Surrey, GB
- Li-Yi Wei
Adobe Inc. – San José, US
- Gordon Wetzstein
Stanford University, US
- Erroll Wood
Microsoft Research –
Cambridge, GB
- Feng Xu
Tsinghua University Beijing, CN