DAGSTUHL
REPORTS

Volume 15, Issue 2, February 2025

 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
https://www.dagstuhl.de/dagpub/2192-5283

Publication date
October, 2025

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 `https://dnb.d-nb.de`.

Digital Object Identifier: 10.4230/DagRep.15.2.i

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

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

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

Report from Dagstuhl Seminar 25061

Logic and Neural Networks

Vaishak Belle*[1], Michael Benedikt*[2], Dana Drachsler-Cohen*[3], Daniel Neider*[4], and Tom Yuviler†[5]

1 **University of Edinburgh, GB.** vaishak@ed.ac.uk
2 **University of Oxford, GB.** michael.benedikt@cs.ox.ac.uk
3 **Technion, IL.** ddana@ee.technion.ac.il
4 **TU Dortmund University, DE.** daniel.neider@tu-dortmund.de
5 **Technion – Haifa, IL.** tom.yuviler@campus.technion.ac.il

Abstract

Logic and learning are central to Computer Science, and in particular to AI-related research. Already Alan Turing envisioned in his 1950 "Computing Machinery and Intelligence" paper a combination of statistical (ab initio) machine learning and an "unemotional" symbolic language such as logic. The combination of logic and learning has received new impetus from the spectacular success of deep learning systems.

This report documents the program and the outcomes of Dagstuhl Seminar 25061 "Logic and Neural Networks". The goal of this Dagstuhl Seminar was to bring together researchers from various communities related to utilizing logical constraints in deep learning and to create bridges between them via the exchange of ideas. The seminar focused on a set of interrelated topics: enforcement of constraints on neural networks, verifying logical constraints on neural networks, training using logic to supplement traditional supervision, and explanation and approximation via logic. This Dagstuhl Seminar aimed not at studying these areas as separate components, but in exploring common techniques among them as well as connections to other communities in machine learning that share the same broad goals.

The seminar format consisted of long and short talks, as well as breakout sessions. We summarize the motivations and proceedings of the seminar, and report on the abstracts of the talks and the results of the breakout sessions.

Seminar February 02–07, 2025 – https://www.dagstuhl.de/25061

2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Computing methodologies → Machine learning approaches; Computing methodologies → Logical and relational learning; Theory of computation → Constraint and logic programming; Theory of computation → Models of learning; General and reference → Surveys and overviews; Theory of computation → Modal and temporal logics

Keywords and phrases machine learning, learning theory, logic, computational complexity, databases, verification, safety

Digital Object Identifier 10.4230/DagRep.15.2.1

* Editor / Organizer
† Editorial Assistant / Collector

Logic and Neural Networks, *Dagstuhl Reports*, Vol. 15, Issue 2, pp. 1–20
Editors: Vaishak Belle, Michael Benedikt, Dana Drachsler-Cohen, Daniel Neider, and Tom Yuviler
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Vaishak Belle (University of Edinburgh, GB, vaishak@ed.ac.uk)
Michael Benedikt (University of Oxford, GB, michael.benedikt@cs.ox.ac.uk)
Dana Drachsler-Cohen (Technion, IL, ddana@ee.technion.ac.il)
Daniel Neider (TU Dortmund University, DE, daniel.neider@tu-dortmund.de)

Motivation

Logic and learning are central to Computer Science, and in particular to AI-related research. Already Alan Turing envisioned in his 1950 "Computing Machinery and Intelligence" paper [1] a combination of statistical (ab initio) machine learning and an "unemotional" symbolic language such as logic. The combination of logic and learning has received new impetus from the spectacular success of deep learning systems. As part of these developments, several key roles for logical rules have been identified: As a means of expressing safety properties that a network should satisfy; As a way of providing "weak supervision", that can be utilized in training, to augment or to substitute for direct supervision; As a means of explaining properties of networks, or explanations of the decisions produced by them. With the identification of these roles, a number of core challenges have arisen: Verifying logic-based properties of networks, Enforcing logic-based properties during training; Utilizing logic-based properties in tandem with traditional supervision within learning to train networks; and Producing logic-based explanations of neural network outcomes. Clearly, these challenges have significant synergy between them. The goal of this seminar was to bring together researchers from various communities related to utilizing constraints in deep learning, and to create bridges between them via the exchange of ideas.

Design of the Seminar

The seminar focused on a set of interrelated topics connected to logic and neural networks:

- **Verifying logical constraints on neural networks.** Despite being successful in various tasks, neural networks have also been shown to be susceptible to various attacks (e.g., adversarial attacks [2]) or prone to biased decisions (e.g., in Amazon's systems[1]). To understand the resilience of networks to these phenomena, it is crucial to prove that networks satisfy *safety properties*, such as local robustness and fairness. These are captured via *logical constraints*, defined on specific inputs in a given dataset (e.g., local robustness) or universally on any input (e.g., fairness and global robustness). Many works have proposed verification systems for these properties [3], typically leveraging constraint solvers [4, 5] or static analysis [6, 7]. Constraints can derive from a number of motivations: security/safety, fairness, or interpretability. Despite the active research on verifying these properties, existing approaches still do not scale to very deep networks, which are ubiquitous in practice. We believe it is viable to understand how to push forward the analysis capabilities to use them for large and deep networks. This will have an impact both for academy and industry, since it will increase the users' trust in practical neural network-based systems.

[1] e.g., https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

- **Enforcement of constraints on neural networks.** Logical rules can represent important safety properties and prior knowledge into the training of neural networks. For example, in a manufacturing setting, we may wish to encode that an actuator for a robotic arm does not exceed some threshold (e.g., causing the arm to move at a hazardous speed). Another example is a self-driving car, where a controller should be known to operate within a predefined set of constraints (e.g., the car should always stop completely when facing a human). In such safety-critical domains, machine learning solutions must guarantee to operate within distinct boundaries that are specified by logical, arithmetic, and geometric rules. Techniques include specialized loss functions [8], which can be augmented with additional layers within a neural architecture. These approaches compile the constraints into the loss function of the training algorithm, by quantifying the extent to which the output of the network violates the constraints. This is appealing as logical constraints are easy to elicit from people. However, the solution outputted by the network is designed to minimize the loss function – which combines both data and constraints – rather than to guarantee the satisfaction of the domain constraints. So this is an important open problem.
- **Training using logic and traditional supervision.** A major of impetus for a synthesis of logic and learning relates to paucity of supervision. In many regimes explicit supervision is extremely limited, and synthetic data generation may be infeasible. A promising approach to augment supervision is via the use of external knowledge. The approach has been used in domains as distinct as scene recognition [9] and parsing [10]. Approaches that integrate constraint-based supervision with traditional supervision have arisen simultaneously in many areas of artificial intelligence. While the focus of our seminar is constraints expressed in general-purpose logics, we look for connections with constraint-based approaches to learning from other areas, such as physics.
- **Explaining neural networks via logic.** A critical issue with black box models, particularly neural networks, is understanding their decision boundaries. An important strategy employed in recent years involves attempting to extract decision trees, logical rules, and other deterministic machines from these neural networks [11, 12, 13, 14]. This can be seen as a strategy for post-hoc explanation [15]. Most approaches for rule extraction use template-based approaches to explore patterns in pre-trained models, with a focus on characteristics and properties of entities such as people, places, or things. However, template-based approaches do show sensitivity to template formulation, highlighting the need to explore alternative strategies to probe pre-trained models. They are often based on a combination of techniques from Bayesian Structure Learning, Inductive Logic Programming [16], and Distillation [15]. Explanation and approximation via logics have also arisen in Graph Neural Networks [17]. An interesting phenomenon is that one of the languages used for explanation is Datalog, which is also prominent in the verification community. The ability to approximate networks by logics is closely-related to attempts to understand the expressiveness of neural approaches in terms of logics [18, 19].

Summary of Seminar Activities

The seminar was attended by 38 researchers across various communities including logic, formal verification, machine learning, deep learning, program synthesis, graph neural networks, expressiveness, explainability, theorem proving, neural-symbolic learning, and databases. The seminar participants included senior and junior researchers, including graduate students, postdoctoral researchers, faculty members, and industry experts. The seminar was conducted through talks and breakout sessions, with breaks for discussion between the attendees. Overall, there were 19 talks, and two main breakout sessions. The talks included a range of presentations on recent advances in the interrelated fields of logic and neural networks, as previously discussed. Some talks also provided broader overviews of related areas, such as formal verification of neural networks and program synthesis. The first breakout session was divided into four groups based on the participants' main areas of research: verification, expressivity, explainability, and learning with background knowledge and constraints. Each group discussed several topic-specific questions: (1) the open challenges, (2) the value proposition, (3) potential "killer applications" or teaching curricula, and (4) drafting a concise manifesto. The second breakout session was divided into three groups (based on participants' choices), each focused on integrating interrelated topics: (1) verification and constraints, (2) explainability, expressiveness, and constraints, and (3) verification and explainability. Each group examined several issues concerning the interplay of these areas, including: (1) prior work, (2) open challenges, (3) real-world motivations and applications, and (4) short- and long-term project ideas.

Conclusion

We consider the seminar a success and believe it achieved several goals that will help strengthen connections among the fields of neural-network verification, logic, explainability, and expressivity. These include: (1) fostering links among the participating researchers, (2) generating a set of open challenges, goals, and future research directions, and (3) providing a more unified view of current approaches to these interrelated topics. We also hope the seminar will catalyze the further development of benchmarks for applying logic in neural networks. Finally, the seminar's format – featuring talks, ample time for discussion, and breakout sessions – received positive feedback from participants.

References

1. A. M. Turing: Computing machinery and intelligence. In Mind, vol. LIX (1950)
2. Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian J. Goodfellow, Rob Fergus: Intriguing properties of neural networks. In ICLR (2014).
3. Linyi Li, Tao Xie, Bo Li. SoK: Certified robustness for deep neural networks. In IEEE Symposium on Security and Privacy (2023).
4. Guy Katz, Clark W. Barrett, David L. Dill, Kyle Julian, Mykel J. Kochenderfer: Reluplex: a calculus for reasoning about deep neural networks. In Formal Methods Syst. Des. (2022).
5. Vincent Tjeng, Kai Yuanqing Xiao, Russ Tedrake: Evaluating robustness of neural networks with mixed integer programming. In ICLR (2019).
6. Timon Gehr, Matthew Mirman, Dana Drachsler-Cohen, Petar Tsankov, Swarat Chaudhuri, Martin T. Vechev: AI2: safety and robustness certification of neural networks with abstract interpretation. In IEEE Symposium on Security and Privacy (2018).
7. Huan Zhang, Tsui-Wei Weng, Pin-Yu Chen, Cho-Jui Hsieh, Luca Daniel: Efficient neural network robustness certification with general activation functions. In NeurIPS (2018).
8. Jingyi Xu, Zilu Zhang, Tal Friedman, Yitao Liang, Guy Van den Broeck: A semantic loss function for deep learning with symbolic knowledge. In ICML (2018).
9. Eleonora Giunchiglia, Mihaela Catalina Stoian, Salman Khan, Fabio Cuzzolin, Thomas Lukasiewicz: ROAD-R: the autonomous driving dataset with logical requirements. In Mach. Learn. (2023).

10 Chen Liang, Jonathan Berant, Quoc V. Le, Ken Forbus, Ni Lao. Neural symbolic machines: Learning semantic parsers on freebase with weak supervision. In ACL (2017).

11 Jan Ruben Zilke, Eneldo Loza Mencía, Frederik Janssen: Deepred–rule extraction from deep neural networks. In DS (2016).

12 Robert Andrews, Joachim Diederich, Alan B Tickle: Survey and critique of techniques for extracting rules from trained artificial neural networks. In Knowledge-based systems (1995).

13 Tameru Hailesilassie: Rule extraction algorithm for deep neural networks: A review. In CoRR abs/1610.05267 (2016).

14 Hiroshi Tsukimoto: Extracting rules from trained neural networks. In IEEE Transactions on Neural networks (2000).

15 Vaishak Belle, Ioannis Papantonis. Principles and practice of explainable machine learning. In Frontiers in big Data (2021).

16 Stephen Muggleton, Luc De Raedt, David Poole, Ivan Bratko, Peter Flach, Katsumi Inoue, Ashwin Srinivasan: ILP turns 20. In Machine learning (2012).

17 David Tena Cucala, Bernardo Cuenca Grau, Boris Motik, and Egor V. Kostylev: On the correspondence between monotonic max-sum gnns and datalog. In KR (2023).

18 Floris Geerts, Juan L. Reutter: Expressiveness and approximation properties of graph neural networks. In ICLR (2022).

19 David Chiang, Peter Cholak, Anand Pillay: Tighter bounds on the expressivity of transformer encoders. In ICML (2023).

2 Table of Contents

3 Overview of Talks

3.1 Learning Symmetric Rules with SATNet

Hongseok Yang (KAIST – Daejeon, KR, hongseok00@gmail.com)

Joint work of Hongseok Yang, Sangho Lim, Eungyeol Oh
Main reference Sangho Lim, Eun-Gyeol Oh, Hongseok Yang: "Learning Symmetric Rules with SATNet", in Proc. of the Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 – December 9, 2022, 2022.
URL http://papers.nips.cc/paper_files/paper/2022/hash/5642b9811a9ac5281be1cc84c275f251-Abstract-Conference.html

SATNet is a differentiable constraint solver with a custom backpropagation algorithm, which can be used as a layer in a deep-learning system. It is a promising proposal for bridging deep learning and logical reasoning. In fact, SATNet has been successfully applied to learn, among others, the rules of a complex logical puzzle, such as Sudoku, just from input and output pairs where inputs are given as images. In this talk, I explain our work on improving the learning of SATNet by exploiting symmetries in the target rules of a given but unknown logical puzzle or more generally a logical formula. I present SymSATNet, a variant of SATNet that translates the given symmetries of the target rules to a condition on the parameters of SATNet and requires that the parameters should have a particular parametric form that guarantees the condition. The requirement dramatically reduces the number of parameters to learn for the rules with enough symmetries, and makes the parameter learning of SymSATNet much easier than that of SATNet. I also describe a technique for automatically discovering symmetries of the target rules from examples. Our experiments with Sudoku and Rubik's cube show the substantial improvement of SymSATNet over the baseline SATNet.

3.2 Query Languages for Machine Learning Models

Pablo Barcelo (PUC – Santiago de Chile, CL, pbarcelo@uc.cl)

Emerging challenges in machine learning (ML), such as explainability and verification, underscore the growing need for declarative query languages that enable users to extract relevant information from ML models and adapt it to diverse application-specific requirements. These query languages offer several advantages: they provide flexibility in information extraction, establish clear syntax and semantics for queries, and pave the way for query optimization. In this talk, we survey two recent proposals for query languages tailored to ML models – one designed for discrete classification models and another for real-valued models. We demonstrate how these languages can express meaningful queries over ML models, and we analyze their expressiveness and evaluation complexity. Our goal is to foster a productive discussion on advancing the development of practical query languages for ML models that can be effectively applied across a wide range of scenarios.

3.3 How to make logics neurosymbolic

Luc De Raedt (KU Leuven, BE, luc.deraedt@kuleuven.be)

Neurosymbolic AI (NeSy) is regarded as the third wave in AI. It aims at combining knowledge representation and reasoning with neural networks. Numerous approaches to NeSy are being developed and there exists an 'alphabet-soup' of different systems, whose relationships are often unclear. I discuss the state-of-the art in NeSy and argue that there are many similarities with statistical relational AI (StarAI). Taking inspiring from StarAI, and exploiting these similarities, I argue that Neurosymbolic AI = Logic + Probability + Neural Networks. I also provide a recipe for developing NeSy approaches: start from a logic, add a probabilistic interpretation, and then turn neural networks into 'neural predicates'. Probability is interpreted broadly here, and is necessary to provide a quantitative and differentiable component to the logic. At the semantic and the computation level, one can then combine logical circuits (aka proof structures) labelled with probability, and neural networks in computation graphs. I illustrate the recipe with NeSy systems such as DeepProbLog, a deep probabilistic extension of Prolog, and DeepStochLog, a neural network extension of stochastic definite clause grammars (or stochastic logic programs).

3.4 Bridging Generalization and Expressivity of Graph Neural Networks

Floris Geerts (University of Antwerp, BE, floris.geerts@uantwerp.be)

Joint work of Shouheng Li, Floris Geerts, Dongwoo Kim, Qing Wang
Main reference Shouheng Li, Floris Geerts, Dongwoo Kim, Qing Wang: "Towards Bridging Generalization and Expressivity of Graph Neural Networks", in Proc. of the The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025, OpenReview.net, 2025.
URL https://openreview.net/forum?id=BOQpRtI4F5

The expressive power of graph neural networks (GNNs) has been widely analysed through their connection to the 1-dimensional Weisfeiler–Leman (1-WL) algorithm, a key tool for addressing the graph isomorphism problem. While this link has deepened our understanding of how GNNs represent complex structures, it provides limited insight into their generalisation – specifically, their ability to accurately predict on unseen data. In this talk, we delve into the relationship between GNNs' expressive power and their generalisation capabilities, offering a perspective that bridges these two critical aspects of GNN performance.

3.5 Formal Verification of Machine Learning with the Industry: The Journey so Far, And the Future Ahead

Julien Girard-Satabin (CEA de Saclay – Gif-sur-Yvette, FR, julien.girard2@cea.fr)

Joint work of Julien Girard-Satabin, Augustin Lemesle, Julien Lehmann, Tristan Le Gall
Main reference Augustin Lemesle, Julien Lehmann, Tristan Le Gall: "Neural Network Verification with PyRAT", CoRR, Vol. abs/2410.23903, 2024.
URL https://doi.org/10.48550/ARXIV.2410.23903

Since the third AI revolution in 2012, industry displayed a keen interest in the newfound capabilities of machine learning. However, in the field of critical systems, existing regulations and practices require some degree of formal specification (and verification). Furthermore, machine learning specification is implicitly defined by hyperparameters that are impossible to formalise (the dataset, the architecture, the objective function, the intended goal). To address those newfound challenges and fulfill its mission to support industrial actors, the French Atomic Energy Commission develop and maintain several tools for the specification and verification of machine learning systems. For seven years, those tools were applied in industrial settings, in national and international projects. Through this presentation mixing science and technical retrospective, we present the successes, the limitations and potential future paths for formal verification informed by the needs of the French industry.

3.6 Learning with Constraints: Fuzzy Methods

Eleonora Giunchiglia (Imperial College London, GB, e.giunchiglia@imperial.ac.uk)

Joint work of Eleonora Giunchiglia, Mihaela Catalina Stoian, Thomas Lukasiewicz
Main reference Eleonora Giunchiglia, Mihaela Catalina Stoian, Thomas Lukasiewicz: "Deep Learning with Logical Constraints", in Proc. of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022, pp. 5478–5485, ijcai.org, 2022.
URL https://doi.org/10.24963/IJCAI.2022/767

In this first segment of the tutorial I discuss methods based on fuzzy logic for learning with constraints. In this talk, I first provide an overview of the learning tasks where logical constraints can play a fundamental role. Then I introduce the most commonly used triangular norms, i.e., Gödel, Product and Lukasiewicz, describing their properties. This be followed by the introduction of "Logic Tensor Network" (LTN), which is one of the most famous methods to integrate constraints in neural networks' loss functions and "Coherent-by-Construction Network" (CCN+), a method to integrate constraints in a neural layer. Both methods are based on triangular norms. After this overview, I also discuss how – thanks to the versatility of fuzzy logic – we can now build neural layers integrating constraints as expressive as disjunctions over linear inequalities, which hence model non-convex and disconnected spaces. I conclude the talk with a discussion with the pros and cons of using fuzzy methods in learning with constraints.

3.7 Neural Continuous-Time Supermartingale Certificates

Anna Lukina (TU Delft, NL, A.Lukina@tudelft.nl)

Joint work of Grigory Neustroev, Mirco Giacobbe, Anna Lukina
Main reference Grigory Neustroev, Mirco Giacobbe, Anna Lukina: "Neural Continuous-Time Supermartingale Certificates", in Proc. of the AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 – March 4, 2025, Philadelphia, PA, USA, pp. 27538–27546, AAAI Press, 2025.
URL https://doi.org/10.1609/AAAI.V39I26.34966

We introduce for the first time a neural-certificate framework for continuous-time stochastic dynamical systems. Autonomous learning systems in the physical world demand continuous-time reasoning, yet existing learnable certificates for probabilistic verification assume discretization of the time continuum. Inspired by the success of training neural Lyapunov certificates for deterministic continuous-time systems and neural supermartingale certificates for stochastic discrete-time systems, we propose a framework that bridges the gap between continuous-time and probabilistic neural certification for dynamical systems under complex requirements. Our method combines machine learning and symbolic reasoning to produce formally certified bounds on the probabilities that a nonlinear system satisfies specifications of reachability, avoidance, and persistence. We present both the theoretical justification and the algorithmic implementation of our framework and showcase its efficacy on popular benchmarks.

3.8 Challenges for the Certification of AI in Railway Systems

Pierre-Jean Meyer (Gustave Eiffel University – Villeneuve d'Ascq, FR, pierre-jean.meyer@univ-eiffel.fr)

The trend of AI and desire to develop autonomous rail vehicles has led to a surge of interest for the use of AI in the railway field, including in safety-critical functions. Traditionally in the railway field, formal methods have been strongly recommended for the certification of safety-related components, but currently applied approaches cannot be properly adapted for the certification of AI functions. This talk gives a brief overview of current and desired applications of AI in railway field, as well as the main identified challenges for the use of formal verification to certify the good behaviors of AI functions within safety-related modules in autonomous trains: primarily the computational complexity and the definition of formal specifications.

3.9 Distinguished In Uniform: Self Attention Vs. Virtual Nodes

Martin Ritzert (Universität Göttingen, DE, ritzert@informatik.uni-goettingen.de)

Joint work of Eran Rosenbluth, Jan Tönshoff, Martin Ritzert, Berke Kisin, Martin Grohe
Main reference Eran Rosenbluth, Jan Tönshoff, Martin Ritzert, Berke Kisin, Martin Grohe: Distinguished In Uniform: Self-Attention Vs. Virtual Nodes. In ICLR (2024).
URL https://openreview.net/forum?id=AcSChDWL6V

Graph Transformers (GTs) such as SAN and GPS are graph processing models that combine Message-Passing GNNs (MPGNNs) with global Self-Attention. They were shown to be universal function approximators, with two reservations: 1. The initial node features must be augmented with certain positional encodings. 2. The approximation is non-uniform: Graphs of different sizes may require a different approximating network. We first clarify that this form of universality is not unique to GTs: Using the same positional encodings, also pure MPGNNs and even 2-layer MLPs are non-uniform universal approximators. We then consider uniform expressivity: The target function is to be approximated by a single network for graphs of all sizes. There, we compare GTs to the more efficient MPGNN + Virtual Node architecture. The essential difference between the two model definitions is in their global computation method – Self-Attention Vs. Virtual Node. We prove that none of the models is a uniform-universal approximator, before proving our main result: Neither model's uniform expressivity subsumes the other's. We demonstrate the theory with experiments on synthetic data. We further augment our study with real-world datasets, observing mixed results which indicate no clear ranking in practice as well.

3.10 How Can Formal Methods Benefit Large Language Models

Gagandeep Singh (University of Illinois – Urbana-Champaign, US, ggnds@illinois.edu)

Joint work of Isha Chaudhary, Qian Hu, Manoj Kumar, Morteza Ziyadi, Rahul Gupta, Gagandeep Singh
Main reference Isha Chaudhary, Qian Hu, Manoj Kumar, Morteza Ziyadi, Rahul Gupta, Gagandeep Singh: "Certifying Counterfactual Bias in LLMs", in Proc. of the The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025, OpenReview.net, 2025.
URL https://openreview.net/forum?id=HQHnhVQznF

Despite impressive performance, state-of-the-art Large Language Models (LLMs) often hallucinate, produce toxic responses, and leak sensitive information. While increasing model sizes, using more training data, compute resources, and prompt engineering have some marginal impact on LLM behavior, these ad-hoc methods do not solve the core problems. Further. These solutions are unsustainable due to their huge environmental impact. In this talk, I discuss how formal methods can be leveraged to develop principled and systematic approaches to improve LLM performance and alignment, offering a path forward that is both effective and sustainable.

3.11 Program Synthesis Present and Future

Armando Solar-Lezama (MIT – Cambridge, US, asolar@csail.mit.edu)

Large Language Models have transformed the landscape of program synthesis, enabling us to solve previously intractable problems and opening up new applications. In this talk I give a high-level summary of the current state of the art in program synthesis and describe some of the open problems and opportunities in the field.

3.12 Refining Deep Generative Modelling using Background Knowledge

Mihaela Stoian (University of Oxford, GB, mihaela.stoian@cs.ox.ac.uk)

Joint work of Mihaela C. Stoian, Salijona Dyrmishi, Maxime Cordy, Thomas Lukasiewicz, Eleonora Giunchiglia
Main reference Mihaela C. Stoian, Salijona Dyrmishi, Maxime Cordy, Thomas Lukasiewicz, Eleonora Giunchiglia: "How Realistic Is Your Synthetic Data? Constraining Deep Generative Models for Tabular Data", in Proc. of the The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024, OpenReview.net, 2024.
URL https://openreview.net/forum?id=tBROYsEz9G

Synthesising realistic tabular data often relies on deep generative models. However, these models fail to account for inherent relationships between features, encoded as background knowledge, which synthetic samples must satisfy to be deemed realistic. Existing methods handle non-compliant samples by discarding them, leading to potentially indefinite inference times. In this talk, I present a novel approach that embeds a constraint layer into the topology of deep generative models to account for the relationships between the features. This layer automatically incorporates background knowledge and ensures compliance with these constraints during both training and inference. I first present our method for handling linear constraints and then discuss its extension to support quantifier-free linear real arithmetic constraints. Experimental results show that our layer significantly improves the machine learning efficacy of deep generative models without hindering sample generation times. This framework is part of our broader goal of bringing neuro-symbolic AI onto the stage of real-world applications.

3.13 Expressive Power of Graph Neural Networks via Datalog

David Tena Cucala (Royal Holloway, University of London, GB, David.TenaCucala@rhul.ac.uk)

This talk discusses recent results on the expressive power of Graph Neural Networks (GNNs) operating on relational datasets. We consider two sub-classes of GNNs: monotonic GNNs and max GNNs, and then we identify Datalog fragments or extensions that realize the same transformations as these GNNs. Monotonic GNNs are GNNs subject to restrictions ensuring that their behaviour is monotonic under homomorphisms applied to their input. Max GNNs are subject to the restriction that they use the max function to aggregate information. Finally, we illustrate some applications of these results, in the areas of GNN verification and explanation of predictions.

3.14 Static Analysis Methods for Neural Networks

Caterina Urban (INRIA and ENS Paris, FR, caterina.urban@inria.fr)

Joint work of Caterina Urban, Maria Christakis, Valentin Wüstholz, Fuyuan Zhang
Main reference Caterina Urban, Maria Christakis, Valentin Wüstholz, Fuyuan Zhang: "Perfectly parallel fairness certification of neural networks", Proc. ACM Program. Lang., Vol. 4(OOPSLA), pp. 185:1–185:30, 2020.
URL https://doi.org/10.1145/3428253

Formal methods provide rigorous guarantees of correctness for both hardware and software systems. Their use is well established in industry, notably to certify safety of critical applications subject to stringent certification processes. With the rising prominence of machine learning, the integration of machine-learned components into critical systems presents novel challenges for the soundness, precision, and scalability of formal methods. This talk serves as an introduction to formal methods tailed for machine learning software, with a focus on static analysis methods for neural networks. We present several verification approaches, highlighting their strengths and limitations, through the lens of different (hyper)safety properties. A neural network surrogate from a real-world avionics use case serves as a running example. We additionally survey the application of these verification approaches towards the additional goal of enhancing machine learning explainability. We conclude with perspectives on possible future research directions in this rapidly evolving field.

3.15 From Learning with Constraints to Partial Label Learning

Zsolt Zombori (Alfréd Rényi Institute of Mathematics – Budapest, HU, zombori@renyi.hu)

Joint work of Zsolt Zombori, Agapi Rissaki, Kristóf Szabó, Wolfgang Gatterbauer, Michael Benedikt
Main reference Zsolt Zombori, Agapi Rissaki, Kristóf Szabó, Wolfgang Gatterbauer, Michael Benedikt: "Towards Unbiased Exploration in Partial Label Learning", CoRR, Vol. abs/2307.00465, 2023.
URL https://doi.org/10.48550/ARXIV.2307.00465

In numerous learning setups, some background knowledge is available in the form of logical constraints. Such constraints can be useful both for increasing the safety of the trained models and for alleviating data shortage by making learning more effective. In this talk we review different types of constraints and how they can possibly be incorporated into the learning or inference process. We also identify a bias phenomenon that occurs during gradient descent based optimisation with constraints, preventing proper exploration of alternative options and making the dynamics of gradient descent overly sensitive to initialisation. We introduce a novel loss function that allows for unbiased exploration within the space of alternative outputs.

3.16 Learning with Constraints: Probabilistic Methods

Emile van Krieken (University of Edinburgh, GB, Emile.van.Krieken@ed.ac.uk)

Joint work of Emile van Krieken, Thiviyan Thanapalasingam, Jakub M. Tomczak, Frank van Harmelen, Annette ten Teije
Main reference Emile van Krieken, Thiviyan Thanapalasingam, Jakub M. Tomczak, Frank van Harmelen, Annette ten Teije: "A-NeSI: A Scalable Approximate Method for Probabilistic Neurosymbolic Inference", in Proc. of the Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 – 16, 2023, 2023.
URL http://papers.nips.cc/paper_files/paper/2023/hash/4d9944ab3330fe6af8efb9260aa9f307-Abstract-Conference.html

I discuss probabilistic methods for learning with constraints. First, I recap practical issues with fuzzy methods. Then, I introduce the weighted model count (WMC), the central equation underlying probabilistic methods for integrating constraints. The WMC gives many theoretical guarantees. With the WMC at hand, I describe a popular constraint loss method called "Semantic Loss", and a constraint layer called "Semantic Probabilistic Layers". This part ends with a comparison of the strengths and weaknesses of probabilistic and fuzzy methods. After this introduction to the core methods, I describe several issues with constraint losses, starting with Reasoning Shortcuts. This is the phenomenon that models may completely minimise the constraint training loss without learning underlying concepts. I also discuss issues with a conditional independence assumption that is frequently taken in practical setups. I end with a brief introduction of state-of-the-art methods for tackling these issues, and a recap of the methods discussed in this two-part tutorial.

4 Breakout Sessions

4.1 Verification

Vaishak Belle (University of Edinburgh, GB, vaishak@ed.ac.uk)
Michael Benedikt (University of Oxford, GB, michael.benedikt@cs.ox.ac.uk)
Dana Drachsler-Cohen (Technion, IL, ddana@ee.technion.ac.il)
Daniel Neider (TU Dortmund University, DE, daniel.neider@tu-dortmund.de)

Joint work of All participants

Formal verification of ML is currently overly focusing on a very specific set of properties, whose real-world applicability may not be fully correlated with the amount of work poured into it. The community must extend towards the ML community and regulators to provide expressive, sound tools that help better characterize complex systems (for instance, multiple NNs or complex constraints on data) with expressive languages and principled compilation toward provers. For this endeavor to be realized, languages should be accessible to non-experts (possibly through constrained means). Furthermore, verifiers should scale to realistic settings, and creative ways to devise specifications should be pursued, for instance by synthesizing properties.

4.2 Expressivity

Vaishak Belle (University of Edinburgh, GB, vaishak@ed.ac.uk)
Michael Benedikt (University of Oxford, GB, michael.benedikt@cs.ox.ac.uk)
Dana Drachsler-Cohen (Technion, IL, ddana@ee.technion.ac.il)
Daniel Neider (TU Dortmund University, DE, daniel.neider@tu-dortmund.de)

Joint work of All participants

Research on expressivity in advanced ML architectures highlights several open challenges, such as how Transformers handle compositions and whether components like positional encodings are truly necessary. The community must develop principled logical frameworks that clarify which expressions are learnable, while balancing "succinctness" and expressivity so that models remain trainable in realistic settings. Logical upper and lower bounds can guide the design of new architectures and help prevent unintended behaviors (e.g., through constrained losses or temporal constraints). Success stories such as the Weisfeiler-Lehman (WL) approach in graph learning show the value of bridging logical theory and ML practice, though some models (like k-WL) have proven impractical. Looking ahead, we should refine these analyses for GNNs, consider how different architectural features shape learning, and pursue sound yet accessible methods that integrate logic and machine learning across diverse applications.

4.3 Explainability

Vaishak Belle (University of Edinburgh, GB, vaishak@ed.ac.uk)
Michael Benedikt (University of Oxford, GB, michael.benedikt@cs.ox.ac.uk)
Dana Drachsler-Cohen (Technion, IL, ddana@ee.technion.ac.il)
Daniel Neider (TU Dortmund University, DE, daniel.neider@tu-dortmund.de)

Joint work of All participants

Research on explainability in ML highlights challenging trade-offs between model performance and interpretability, particularly in high-stakes domains where trust and transparency are paramount. While methods such as SHAP or LIME offer partial insights, the community still grapples with fundamental questions about how to align expert understanding with possibly less accurate yet more transparent models. Practical benefits extend beyond improved decision-making: interpretable systems can foster scientific discovery by exposing the reasoning behind model predictions, enabling knowledge transfer across tasks, and ensuring that ethical constraints are thoroughly verified. Ultimately, progress in explainability hinges on identifying scenarios where transparent models demonstrably outperform black-box approaches, attracting broader funding and community engagement, and integrating reverse-engineerable explanations that help pinpoint out-of-distribution cases and other critical failures.

4.4 Learning with Background Knowledge & Constraints

Vaishak Belle (University of Edinburgh, GB, vaishak@ed.ac.uk)
Michael Benedikt (University of Oxford, GB, michael.benedikt@cs.ox.ac.uk)
Dana Drachsler-Cohen (Technion, IL, ddana@ee.technion.ac.il)
Daniel Neider (TU Dortmund University, DE, daniel.neider@tu-dortmund.de)

Joint work of All participants

There are several key challenges and considerations in integrating constraints into AI systems, particularly in enhancing performance across various metrics such as safety, data efficiency, accuracy, model size, and generalizability. One of the main challenges identified is demonstrating measurable improvements in real-world scenarios. The discussion also emphasized the need for parametric synthetic datasets with controllable properties, particularly for out-of-distribution (OOD) testing. Additionally, the topic of constraint discovery is highlighted, exploring how constraints can be learned and analyzed in terms of their expressivity, complexity, and geometric properties. The value proposition centers on the advantages of incorporating constraints into AI models beyond just improving accuracy. These advantages include ensuring safer AI decisions, reducing the need for labeled data, enabling more compact models, and improving robustness. The discussion extends beyond traditional models to generative AI, emphasizing that constraints should not only enforce syntactic correctness but also contribute to semantic understanding. We also outline potential teaching material for conveying these neuro-symbolic (NeSy) concepts. The curriculum would begin with a general motivation for NeSy, explaining the complementary strengths of symbolic and statistical approaches. It would introduce key ingredients, including logic (e.g., knowledge graphs, description logics, and logic programming), probabilistic methods, fuzzy logic, neural predicates, and knowledge compilation. The discussion would then cover how these elements integrate into different architectures, addressing aspects such as layers, loss functions, and predicate grounding. A key theme underlying NeSy is encapsulated in the phrase "Why learn what you already know?", suggesting that constraints should guide AI systems by leveraging prior knowledge efficiently.

4.5 Combining Verification and Constraints

Vaishak Belle (University of Edinburgh, GB, vaishak@ed.ac.uk)
Michael Benedikt (University of Oxford, GB, michael.benedikt@cs.ox.ac.uk)
Dana Drachsler-Cohen (Technion, IL, ddana@ee.technion.ac.il)
Daniel Neider (TU Dortmund University, DE, daniel.neider@tu-dortmund.de)

Joint work of All participants

Research on combining constraints and verification in ML underscores the need to ensure that critical requirements remain satisfied, especially when models are treated as black boxes. While constraints can guide the design of more easily verifiable networks – by reducing complexity or limiting nonlinearities – they often need verification to confirm that these properties hold in practice. In turn, verification methods benefit from constraints by narrowing the solution space or allowing for surrogate models that can more efficiently detect potential errors. Challenges persist in communicating across different communities (e.g., security experts operating in black/grey-box settings), devising effective regularizations that maintain performance while improving verifiability, and tackling relational constraints that are notoriously difficult to encode directly. Achieving progress in these areas will require deeper collaboration and possibly new architectures, loss functions, or partitioning strategies that streamline verification while preserving robust performance.

4.6 Combining Explainability, Expressiveness, and Constraints

Vaishak Belle (University of Edinburgh, GB, vaishak@ed.ac.uk)
Michael Benedikt (University of Oxford, GB, michael.benedikt@cs.ox.ac.uk)
Dana Drachsler-Cohen (Technion, IL, ddana@ee.technion.ac.il)
Daniel Neider (TU Dortmund University, DE, daniel.neider@tu-dortmund.de)

Joint work of All participants

Negative results on expressiveness show that logical formulas can be difficult to apply for direct explainability. Traditionally, constraints are motivated by safety rather than explainability, but recent approaches use logical constraints to explain network internals – such as analyzing neuron correlations when given specific images. This strategy could simplify various explainability tasks by leveraging a suitable logical framework, potentially informed by knowledge representation techniques. Several open challenges include determining whether expressive architectures inherently complicate constraint enforcement, identifying parameter sets that minimize constraint violations, and discovering methods to isolate network components corresponding to specific constraints. Addressing these issues could lead to more transparent decision-making grounded in logic-based insights.

4.7 Combining Verification and Explainability

Vaishak Belle (University of Edinburgh, GB, vaishak@ed.ac.uk)
Michael Benedikt (University of Oxford, GB, michael.benedikt@cs.ox.ac.uk)
Dana Drachsler-Cohen (Technion, IL, ddana@ee.technion.ac.il)
Daniel Neider (TU Dortmund University, DE, daniel.neider@tu-dortmund.de)

Joint work of All participants

Formal explanation techniques from classical software – such as SMT solving and UNSAT cores – primarily focus on input/output properties and safety, yet they do not always capture the intricacies of neural networks. Meanwhile, gradient-based attributions in neural networks can be brittle or overly localized, raising questions about how to ensure explanations generalize to unseen instances and how to pinpoint "interventions" that actually shift predictions. Explainable AI (XAI) and verification both rely on abstraction to address these issues: verification uses abstraction to isolate properties that can be formally proven or disproven, whereas explanation refines the model's salient behaviors so users can understand how inputs map to outputs. By capturing properties at a higher level, we obtain amenable properties for explanations, that can be more easily communicated and interpreted. Consequently, joint XAI and verification efforts could devise abstractions that both enable rigorous checks on model correctness and illuminate the model's inner workings. This synergy fosters AI systems that are trustworthy and interpretable, bridging the gap between formal correctness and human-centered understanding.

Participants

- Pablo Barcelo
PUC – Santiago de Chile, CL
- Vaishak Belle
University of Edinburgh, GB
- Michael Benedikt
University of Oxford, GB
- Alexandra Bugariu
MPI-SWS – Kaiserslautern, DE
- Luc De Raedt
KU Leuven, BE
- Dana Drachsler-Cohen
Technion – Haifa, IL
- Sophie Fellenz
RPTU Kaiserslautern-Landau, DE
- Floris Geerts
University of Antwerp, BE
- Julien Girard-Satabin
CEA de Saclay – Gif-sur-Yvette, FR
- Eleonora Giunchiglia
Imperial College London, GB
- Dominik Hintersdorf
TU Darmstadt, DE
- Vivian Holzapfel
Leibniz Universität Hannover, DE
- Omri Isac
The Hebrew University of Jerusalem, IL
- Matthias Lanzinger
TU Wien, AT
- Anji Liu
UCLA, US
- Anna Lukina
TU Delft, NL
- Pierre-Jean Meyer
Gustave Eiffel University – Villeneuve d'Ascq, FR
- Emmanuel Müller
TU Dortmund University, DE
- Daniel Neider
TU Dortmund University, DE
- Ana Ozaki
University of Oslo, NO
- Martin Ritzert
Universität Göttingen, DE
- Patrick Schramowski
DFKI – Darmstadt, DE
- Mahmood Sharif
Tel Aviv University, IL
- Gagandeep Singh
University of Illinois – Urbana-Champaign, US
- Armando Solar-Lezama
MIT – Cambridge, US
- Mihaela Stoian
University of Oxford, GB
- Mislav Stojanovic
TU Dortmund University, DE
- Lukas Struppek
TU Darmstadt, DE
- David Tena Cucala
Royal Holloway, University of London, GB
- Ashish Tiwari
Microsoft Corporation – Redmond, US
- Caterina Urban
INRIA & ENS Paris, FR
- Jan Van den Bussche
Hasselt University, BE
- Frank van Harmelen
VU Amsterdam, NL
- Emile van Krieken
University of Edinburgh, GB
- Jonni Virtema
University of Sheffield, GB
- Hongseok Yang
KAIST – Daejeon, KR
- Tom Yuviler
Technion – Haifa, IL
- Zsolt Zombori
Alfréd Rényi Institute of Mathematics – Budapest, HU

Report from Dagstuhl Seminar 25062

Extended Reality for the Operating Room (XR4OR)

Peter Haddawy*[1], Anja Hennemuth*[2], Ron Kikinis*[3], Gabriel Zachmann*[4], Mario Lorenz†[5], and Anke Reinschlüssel†[6]

1 **Mahidol University – Salaya, TH.** `peter.had@mahidol.ac.th`
2 **Charité – Berlin, DE.** `anja.hennemuth@mevis.fraunhofer.de`
3 **Harvard Medical School – Boston, US.** `kikinis@bwh.harvard.edu`
4 **Universität Bremen, DE.** `zach@cs.uni-bremen.de`
5 **TU Chemnitz, DE.** `mario.lorenz@mb.tu-chemnitz.de`
6 **Universität Konstanz, DE.** `anke.reinschluessel@uni-konstanz.de`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 25062 "Extended Reality for the Operating Room (XR4OR)".

Seminar February 2–7, 2025 – https://www.dagstuhl.de/25062
2012 ACM Subject Classification Applied computing → Health informatics; Human-centered computing → Virtual reality
Keywords and phrases operating room, scientific roadmap
Digital Object Identifier 10.4230/DagRep.15.2.21

1 Executive Summary

Peter Haddawy (Mahidol University – Salaya, TH)
Anja Hennemuth (Charité – Berlin, DE)
Ron Kikinis (Harvard Medical School – Boston, US)
Gabriel Zachmann (Universität Bremen, DE)

The first day started with flash talks where participants could introduce themselves and present their work relevant to the Dagstuhl Seminar. In the afternoon, we discussed the overall outcome of the seminar in the group, a potential research roadmap paper, in order to build a common understanding and engagement with the topic. We also showcased about 10 demos, which several participants brought, so that all participants could get an impression of the current state of the art, and also in order to get everybody in the theme of the seminar. This proved to be very successful, even though participants had to bring all the VR hardware themselves.

Tuesday morning was dedicated to introducing the pain points that surgeons face in their daily work in the operating room to the non-medical attendees, i.e., the virtual reality experts and researchers. We had four talks by four surgeons, each one followed by discussions in order to gain a better understanding. The rest of the day was dedicated to generating a large number of ideas, some of which addressed the pain points, but many of them also were generated very creatively by the participants in break-out sessions using the brain-writing

* Editor / Organizer
† Editorial Assistant / Collector

method. At the end of the day, participants presented their best ideas in the whole group. After dinner, the organizers sat down together in order to cluster all the ideas and create larger umbrella themes.

On Wednesday, the day began by presenting the clusters to all attendees, and by partitioning the group so that each cluster could be taken care of in break-out sessions. The rest of the day was filled by fleshing out the cluster themes and providing more details with the overarching research themes. On Wednesday night, the organizers sat down to create a framework document, to be used on Thursday for starting the writing process.

Thursday was dedicated to collaborative writing. To do so, each theme was further structured by the attendees, some themes were broken down into smaller sub themes, some of them were already narrow enough, so that in the end, we wrote in pairs on each and every topic.

On Friday, we discussed a number of opportunities and possibilities that would allow us to continue this collaboration as a group. To do so, the organizers presented all the possible funding opportunities we were aware of, and we collected some more from the audience. We also discussed a timeline, in order to produce a research roadmap, and we discussed next steps.

Future activities will include writing and publishing a collaborative paper, where each participant will contribute to one section. The paper is to provide perspectives on trends and future research on the seminar's topic, and also it can serve as a guideline or roadmap for researchers who like to enter this field. Further activities will strive to establish and broaden a community around the seminar's topic, by organizing workshops at prestigious VR conferences, such as IEEE VR. Eventually, we will apply for another Dagstuhl Seminar, with a broader participant list, and a somewhat broader topic.

2 Table of Contents

3 Overview of Talks

3.1 Bringing Augmented Reality into the Operating Room: Challenges and Opportunities

Philippe C. Cattin (Universität Basel – Allschwil, CH)

At XR4OR, Philippe C. Cattin presented his research on virtual reality (VR) planning for complex surgical interventions using the Specto Volume Rendering application. He also demonstrated how these planning tools can be transferred into the operating room. While modern surgical loupes reduce eye strain and improve ergonomics during procedures, they also limit the field of view, making it difficult for surgeons to maintain an overview of the entire surgical scene. To address this limitation, integrating augmented reality (AR) into surgical loupes offers a promising solution. AR overlays can provide real-time guidance – such as anatomical landmarks, navigation cues, segmented tumors, or hemodynamic data – without requiring the surgeon to look away from the operative field. By seamlessly blending essential information into the surgeon's line of sight, AR-enhanced loupes can improve situational awareness, support better decision-making, and ultimately contribute to safer and more efficient surgeries. However, no medically certified AR-enhanced surgical loupes are currently available on the market. This needs to be addressed soon by the MedTech Industry. Despite their potential, deploying VR and AR technologies for routine hospital use remains challenging. Beyond their excellent visual experience, AR/VR systems must prove their clinical effectiveness and efficiency. Medical professionals are only likely to adopt these tools broadly if they demonstrably save time in their already busy schedules. One major barrier is the time-consuming and often unreliable setup process. Connectivity issues with AR/VR glasses, frequent system reboots, and general instability make these solutions impractical for daily clinical use – issues that may be tolerated in research settings, but not in routine care. Another obstacle to clinical integration is the lack of standardized interaction interfaces across different AR/VR hardware manufacturers. While laptops universally rely on keyboards and mice, AR/VR devices use a wide array of input methods. Some manufacturers rely on hand controllers with varying designs and button layouts; others use hand tracking with different gesture sets. These inconsistencies hinder seamless adoption and training, further slowing down the integration of AR/VR into clinical workflows.

3.2 Interactive rendering techniques for mixed reality simulation, planning and guidance of surgical operations

Simon Drouin (ETS – Montreal, CA)

Joint work of Andrey Titov, Simon Drouin, Alejandro Olivares Hernandez, Tina Nantenaina, Naomi Catwell
Main reference Andrey Titov, Marta Kersten-Oertel, Simon Drouin: "Contextual Ambient Occlusion: A volumetric rendering technique that supports real-time clipping", Comput. Graph., Vol. 119, p. 103884, 2024.
URL https://doi.org/10.1016/J.CAG.2024.01.011

This presentation highlighted recent advances by my group on the interactive rendering of medical images for simulation, planning and guidance of surgical operations, with a specific focus on mixed reality displays. We have shown how advanced volumetric rendering techniques can help visualize 4D echography images, how segmented volumes can be interactively volume-rendered in virtual reality to improve depiction of the anatomy for surgical planning and how these interactive rendering techniques can be leveraged to simulate surgical procedures for the training of residents. Following the presentation, a demo of our VR interactive volume rendering system was setup for other seminar participants to try.

3.3 XR Interaction in the Operation Room

Tanja Döring (Universität Bremen, DE)

As a researcher in Human-Computer Interaction, the focus of my work is on innovative and multimodal interaction techniques in different application domains. The domain of the operating room (OR), with its phases of training, preoperative, interoperative, and postoperative activities, offers a multitude of possibilities to be supported by interactive technology to make the procedures more efficient and successful, to improve patient safety and well-being as well as the work conditions of surgeons and OR staff. Nowadays, many digital systems already support the work in the OR, such as by providing image data during surgery. However, the interaction with these systems is often cumbersome, as it takes place as a secondary task next to the primary operation task in sterile conditions and with usually busy hands. Thus, solutions for the interactions with the increasing number of systems are needed that fit this specific and safety-critical environment, align well with the main task of the operation, and do not increase workload. A rise of systems that use extended reality (XR) and artificial intelligence (AI) to provide more extensive assistance to surgeons and medical staff will also set new requirements for the interaction with technology during training, preoperative planning, during surgery in the operating room, and postoperative care. A well-balanced design of embodied, multimodal implicit and explicit XR interactions that also consider a trustful and sensible human-AI collaboration is needed. During the Dagstuhl Seminar "Extended Reality for the Operating Room" we fruitfully discussed this highly interdisciplinary, applied, and dynamic field of research from many angles, compiling visions for smooth, intelligent, and effective XR systems of the future. I am interested in working on a design space for XR interaction in the OR, putting together the dimensions that need to be taken into account when designing for this domain, which include the different phases, tasks and goals, the user roles as well as (explicit and implicit) input and output facilities. Especially the integration of implicit interaction, such as through the tracking of eye gaze and biosensing for input, or the inclusion of additional modalities for output, such as haptic feedback, are promising paths to facilitate and enrich XR interactions in the OR. Nowadays, many parallel systems and screens exist that are not integrated. The usability of used XR systems, such as HMDs, still needs to be improved, and interactions lack standards, so they may be challenging to learn and remember. Thus, there is a great demand for standardization and integration and the design of personalized and context-aware, more intelligent interactive assistance systems.

3.4 Gabor Fichtinger: directions for future grants

Gabor Fichtinger (Queen's University – Kingston, CA)

During the Dagstuhl Seminar, I focused on developing research project ideas and exploring directions for future grant proposals. I also contributed to drafting a joint white paper that brought together perspectives from different participants to outline shared research priorities. In addition, I met several researchers whose work intersects with mine, leading to new academic connections that may support future collaboration. The seminar provided time and space for thoughtful exchange, helping identify common interests and potential pathways for joint work in global health.

3.5 Imaging data and preoperative planning as main components for XR in the operating room

Tabea Flügge (Charité – Berlin, DE)

Extended Reality in the operating room may enhance surgical precision, efficiency, and patient outcomes. Key aspects are the accurate registration and display of multimodal imaging data and the transfer of the preoperative planning to the surgical site. XR may provide real-time 3D overlays of patient anatomy, helping surgeons navigate complex procedures with greater accuracy. The holographic visualization of critical structures, like blood vessels, tumors, or nerves, may reduce surgical risks. Transferring preoperative planning to the surgical site allows for real-time navigation of procedures. The surgeon receives intraoperative guidance by superimposing 3D images of the planned surgical result on the region of interest. In reconstructive oral and maxillofacial surgery cases, complex preoperative planning is performed, and surgical guides and patient-individual implants are manufactured. Therefore, adherence to the surgical plan is crucial for avoiding intra- and postoperative complications. Incorporating automated segmentation and analysis of large amounts of patient data further helps predict surgical risks and machine learning models may suggest surgical approaches to improve the precision of a surgical procedure. A second key aspect besides the availability of information is its individual display, e.g., on head-mounted devices, to avoid distraction and work continuously without breaking focus. User profiles for all staff members in the OR may improve the distribution of information in the OR and avoid redundant information.

3.6 LARACROFT: A Collaborative AR Platform for Enhancing Laparoscopic Training in the Operating Theater

Joaquim A. Jorge (University of Lisbon, PT)

Joint work of Joaquim A. Jorge, António Salgueiro, Mona Zavichi, Anderson Maciel, Catarina Moreira, Francisca Leite, Cátia Costa

The LARACROFT platform introduces an augmented reality (AR) framework tailored for remote collaboration in laparoscopic surgery. Designed to operate within the spatial and cognitive constraints of the operating theater, the system supports real-time interaction between local and remote surgeons via a shared AR workspace. It integrates hand tracking, video passthrough, and gaze-directed, foot-triggered input mechanisms to enable effective control without compromising sterility or workflow. In a pilot study, medical students and mentors engaged in standardized surgical tasks, revealing key insights into the impact of AR-supported telecollaboration on motor coordination and communication. Observed limitations, including latency and display resolution, highlighted current technological thresholds in surgical AR. However, participants reported enhanced engagement and situational awareness, pointing toward the platform's potential for scalable deployment in surgical education and support.

3.7 Workflow Considerations for XR in the OR

Luv Kohli (InnerOptic Technology – Hillsborough, US)

XR has the potential to be used in a variety of ways for the OR, such as medical device navigation, visualization of internal structures, identification of cancerous tissue, pre-procedure or intra-procedure planning, information and image display, telemedicine and communication, and many more. Rather than focus here on a specific use case, I offer some thoughts and questions about how we might enable XR's eventual adoption and use in the OR.

There are cases where XR use by patients shows a lot of promise and success (for example, pain distraction, neurological rehabilitation, PTSD). XR technology has advanced considerably in recent years but to my knowledge, it does not yet have widespread adoption in the OR, at least not for physician users. Why?

When researching, developing, and commercializing new technologies for the OR, we face the challenge of how to integrate those technologies into physician workflows. If a new technology takes too much time to set up for each patient, it becomes an opportunity cost with less time available to treat additional patients. If using the technology is too complex, it ends up not being used. For example, some physicians have told me that even though they have access to some sophisticated medical device navigation systems, setup takes long enough that it is not worth the time or complexity.

Also, for an organization to invest in a technology commercialization effort, there must be reasonable evidence that the technology (compared to standard of care or other less disruptive technologies) will improve patient outcomes, enable physicians to do things they otherwise could not do, improve hospital throughput, generate profit for the hospital, or provide some other business benefit.

As researchers and technologists focused on improving patient care with XR, how might we develop XR to enable its use without workflow disruption? For example, what can be done to make setup quick and efficient? How can technology be made robust to failure or emergency situations (e.g., if an emergency arises, can the physician immediately remove a headset and revert to standard of care?)? How can ergonomics be improved, especially for long procedures?

These considerations often end up being procedure-specific because different procedures have different workflow requirements. For complex procedures, a lot of time and mental effort can be spent on communication between physicians and OR staff, so that physicians can see what they need to see. One could imagine developing procedure-specific conversational AI models for a physician to more easily interact via voice, without requiring very precise wording, or without requiring challenging hand gestures or controllers. Procedure-specific inference models may also be able to understand procedure workflow and show "just-in-time" information to the physician, reducing communication overhead.

There are likely many other workflow-related improvements that can be made, and it would be interesting to explore how our research can support physicians and staff in the OR without negatively impacting their workflow.

3.8 Perspective on the Future Use of XR for OR

Mario Lorenz (TU Chemnitz, DE)

XR, in a strict sense, only encompasses the mediation of virtual, computer-generated content for all human senses i.e., visual overlays, haptic or acoustic amplifications. However, XR alone is just one of several technologies (AI, image processing, 3D-printing, diagnostics) driving the progress in medicine and surgery and a focus solely on XR without including those other domains will rather hinder than facilitate XRs uptake in the OR. In general, it is important not to ask ourselves "What can XR do for the surgeon?", but rather "What problems does the surgeons currently have? Where are the limitations of how they can help the patient?" Using this approach, we will discover where XR really brings a benefit and is not just pushed from a technology enthusiastic perspective. We will further discover how XR fits best in to the supporting ecosystem of advanced technologies (AI, image processing, 3D-printing, diagnostics). By facilitating synergies, and win-win-situations, we will foster the transfer into the OR. The following use case shall demonstrate my perspective:

1. **Patient with Face/Head-Trauma.** When the patient arrives in OR the surgeon could get an overlay of the MRI on the patient using AR. This is not a very new use case, but everything one could achieve solely with XR. By broadening the scope of involved technologies, a much more advanced scenario is imaginable. By 3D-scanning the patients head in the OR an AI could calculate how the bone fragments have moved since the MRI was taken and can visualize this updated 3D-model to the surgeon via an AR headset. Another AI, could deduce how the bone fragments have to be put together stepwise, which implants are necessary for this, how standard implants have to be bend and what custom implants need to be 3D-printed. Once this is determined 3 paralyzed processes are imitated (1) the printing of the custom implants; (2) A nurse or a resident surgeon is adapting the shape of the standard implant using a template of the target

shape visualized through and AR display. An AI checks in the background the forming process and provides and audio-visual feedback when the correct shape is reached; (3) The surgeon gets support through the AR glasses by visualising the positions of the bone fragments and the instructions of how they have to be put together using the implants. During the entire surgery the visual guidance is updated based on the tracked deformation of the patient's soft tissue. Another option for this use case for severe trauma or in case of tumour resection is, that the AI calculating the reconstructions steps also uses pictures form the patient before the incident.

2. **Tracking for AR guidance, VR training and automated surgery.** In the future OR, every step of the surgery is tracked. This data can be used to generate training scenarios in VR simulation but also to teach an AI to perform surgery autonomously. Further, it allows to gather the various existing individual surgical techniques and may allow to find the objectively best procedures. In AR supported surgical guidance applications this knowledge is used for generating the AR instructions.
3. **XR for Telesurgery.** In the future telesurgery will not be performed by the surgeon looking at mono- or stereoscopic images of a live feed. Instead, a detailed and complete 3D representation of the patenting will be simulated visually, haptically, auditory and olfactory. This allows the surgeons to perform the surgery on a virtual patient whilst their movements are tracked and translated to according movements of the surgery robot.
4. **Miscellaneous Use Cases.** (1) AI based detection of a sudden vessel rupture, with AR visualization of the rupture spot. (2) AI based in-situ detection of tumours with AR visualisation and guidance. (3) Integrating of health monitor data and other patient data into a single cockpit that is visualized in AR for the surgeon during surgery.

3.9 Experimenting with augmented reality in the operating room: lessons learned

Javier Pascau (Carlos III University of Madrid, ES)

Joint work of Javier Pascau, Mónica García-Sevilla, David García-Mato, Rafael Moreta-Martínez, Alicia Pose-Díez-de-la-Lastra
Main reference Alicia Pose-Díez-de-la-Lastra, Tamas Ungi, David Morton, Gabor Fichtinger, Javier Pascau: "Real-time integration between Microsoft HoloLens 2 and 3D Slicer with demonstration in pedicle screw placement planning", Int. J. Comput. Assist. Radiol. Surg., Vol. 18(11), pp. 2023–2032, 2023.
URL https://doi.org/10.1007/S11548-023-02977-0

In this talk, I presented our experience at Universidad Carlos III de Madrid introducing Augmented Reality (AR) in the Operating Room (OR). Our initial open-source solution combined a 3D-printed reference, a surgical guide, and an application running on a smartphone, offering a versatile approach for visualizing anatomical models augmented over 3D-printed phantoms or directly onto patients during surgery. With this straightforward method, we successfully displayed essential surgical planning information in training sessions for sacral neurostimulation and during craniosynostosis and microtia surgeries. Smartphones and tablets offer numerous advantages, such as affordability, ease of introduction into sterile environments via transparent bags, and the possibility of sharing the AR visualization among multiple users. However, when porting our AR solutions to Head-Mounted Displays (HMDs) like Microsoft Hololens2, the interaction with holograms and the immersive experience greatly improved. Despite current HMD limitations, this approach offers significant benefits for the

OR, including natural hand-gesture interactions and enhanced 3D perception. Our recent work integrating Hololens2 with the 3DSlicer platform has opened new possibilities, combining the advantages of AR visualization with the computational and algorithmic capabilities available in 3DSlicer. These findings were demonstrated after the talk, utilizing a 3D-printed craniosynostosis patient model to visualize AR content both on a tablet and via Hololens2.

3.10 Personal perspective of seminar outcome

Dirk Reiners (University of Central Florida – Orlando, US)

The main outcome for me for this seminar is a much better understanding of the needs of the medical professionals in the OR, in terms of things they can and cannot do, and their requirements of supporting technology like XR. Discussions with doctors and XR researchers led to new ideas for visual representations of relevant data in an integrated context that can be used inside the OR for different kinds of procedures.

3.11 Multimodal Features and User Representation Considerations

Anke V. Reinschluessel (Universität Konstanz, DE)

Main reference Anke Verena Reinschlüssel: "Empowering surgeons – a multimodal interaction approach to improve computer-assisted surgical planning and interventions", 2022.
URL https://doi.org/10.26092/ELIB/1620

Several aspects are important to consider in order to advance the use of extended (XR) or mixed reality (MR) in the operating room (OR). When thinking about MR, we mostly focus on the visual aspects, but MR can also augment or integrate other senses, such as audio or haptic, which can also support a procedure. There have been approaches to use audio or tactile information to guide interventions (e.g., [3] or [5]). Furthermore, also for surgery planning, haptic props have been shown as a promising approach to interacting with image data (e.g., [1, 2] or [4]). Multimodal feedback might also be particularly interesting in a situation where a specialist is joining remotely and is virtually present in the room. Subtle on-site occurrences, like a sound from a device or someone coming into the OR asking a question, might not be picked up and transmitted by the collaboration system by default but distract the team on-site. This could lead to confusion as the reason for the team's attention shift is unclear to the remote specialist. Therefore, it is important to consider what and how it is transmitted in these remote collaboration settings. To create a better sense of co-presence, a system should also not simply represent the position the remote person placed themselves but maybe also add multimodal feedback for the on-site personnel, e.g., heat patches imitating the body heat someone standing there would radiate. Likewise, it might be worthwhile to explore which kind of user representations support remote collaboration best, as humans pick up subtle cues without conscious processing that influence the relationships. For example, our breathing and heart rate are perceived and can synchronise, leading to a feeling of closeness and allowing for an understanding of the other person's state in real settings, which are details that current avatar representations omit. Additionally, as avatar representations can also be a source of distraction, investigating which aspects of the user representation are beneficial for use in a high-stress environment and which add stress. The aspects of this talk were intensely discussed in the context of remote collaboration for surgical contexts, but might also play a role in robotics-assisted surgery or remote robotic surgery. It is worthwhile to investigate how the humans operating them can be present in the OR, and, potentially, the results can also be transferred to robotics representations to increase awareness and collaboration with it.

References

1 Reinschluessel, A., Fischer, R., Schumann, C., Uslar, V., Muender, T., Katzky, U., Kißner, H., Kraft, V., Lampe, M., Lück, T. & Others *Introducing virtual & 3D-printed models for improved collaboration in surgery.* Proceedings Of The 18. Annual Meeting Of The German Society Of Computer-and Robot-Assisted Surgery, Oliver Burgert (Ed.). **18** pp. 253-258 (2019)

2 Reinschluessel, A., Muender, T., Salzmann, D., Döring, T., Malaka, R. & Weyhe, D. *Virtual Reality for Surgical Planning – Evaluation Based on Two Liver Tumor Resections.* Frontiers In Surgery. **9** (2022)

3 Black, D., Lilge, S., Fellmann, C., Reinschluessel, A., Kreuer, L., Nabavi, A., Hahn, H., Kikinis, R. & Burgner-Kahrs, J. *Auditory Display for Telerobotic Transnasal Surgery Using a Continuum Robot.* J. Medical Robotics Res.. **4**, 1950004:1-1950004:14 (2019)

4 Muender, T., Reinschluessel, A., Salzmann, D., Lück, T., Schenk, A., Weyhe, D., Döring, T. & Malaka, R. *Evaluating Soft Organ-Shaped Tangibles for Medical Virtual Reality.* CHI '22, New Orleans, LA, USA, 29 April 2022 – 5 May 2022, Extended Abstracts. pp. 237:1-237:8 (2022)

5 Reinschluessel, A., Cebulla, S., Herrlich, M., Döring, T. & Malaka, R. *Vibro-Band: Supporting Needle Placement for Physicians with Vibrations.* Extended Abstracts Of CHI 2018, Montreal, QC, Canada, April 21-26, 2018. (2018)

3.12 Enhancing Liver Surgery and Hepatic Interventions with AI and XR Technologies

Andrea Schenk (Fraunhofer MEVIS – Bremen, DE)

Artificial intelligence technologies, particularly deep learning, have accelerated and partially automated the analysis of preoperative imaging data such as CT and MRI. This has enabled more widespread creation and use of 3D models for treatment planning, risk analysis and decision support. In particular, liver cancer treatments with multiple therapy options, including open and minimally invasive surgery, have benefited from this development. The addition of information from laboratory values, histology and liver function tests, together with numerical simulations and AI-based prediction of treatment outcomes, has led to simpler, safer and improved therapies. XR tools, particularly augmented reality (AR), provide surgeons with real-time visualizations of complex liver anatomy, enabling greater accuracy during procedures. This enhanced visualization helps to accurately identify critical structures such as blood vessels, bile ducts and distances to tumors, ultimately reducing the risk of complications. The use of XR in surgical training also improves the skills of medical professionals by allowing them to practice techniques in an immersive, risk-free environment. The combination of AI and XR technologies helps ensure that patients receive safer and more effective care in the field of liver surgery, with the potential to significantly improve patient outcomes. In the talk, I presented our own research in this area, including the segmentation of anatomical liver structures and pathologies from CT and MRI data using AI, the calculation of image-based liver function, and therapy simulations for tumor resection and ablation as well as for liver transplantation. In three projects, we extended XR technologies for the transfer of 3D models and planning results into the operating room and developed an extremely fast visualization technology called AVIS, which allows noise-free real-time volume rendering in AR with interactive modification of the transfer function and application of cut planes. An overview of current research topics concluded the presentation.

3.13 Towards task specific (physical) XR solutions

Falko Schmid (Heinrich-Heine-Universitätsklinikum Düsseldorf, DE)

XR is emerging as a valuable tool in the operating room (OR), with the potential to support a range of critical processes and workflows, including immersive communication, visualization of medical imagery, navigation, and ambient communication through technologies such as light and projection. These applications can greatly enhance the effectiveness and efficiency of medical procedures, but to unlock their full potential, several key factors must be addressed. A thorough understanding of the technology itself is essential. Devices like headsets must be carefully evaluated for factors such as battery life, resolution, camera position, field of view, brightness, and sensory capabilities, as these elements significantly impact the quality and feasibility of potential solutions. Additionally, interaction capabilities like gestures, voice control, eye tracking, contextual sensor fusion, and the integration of wearables must be explored to ensure they align with the specific needs of the medical environment. Equally important is a deep understanding of the task at hand. It is crucial to determine whether additional visual information would be beneficial and to evaluate alternative ways to present information, such as through sound, ambient changes, vibration, or external visual cues. It's not just about identifying the type of information required but also understanding how and when it should be delivered to maximize its value. The implementation of natural, intuitive, and robust interactions is key, especially in high-pressure environments like surgery, where ease of use and minimizing distractions are critical. Finally, a comprehensive understanding of clinical processes is necessary to integrate MR effectively into existing workflows. Solutions must be designed to enhance the capabilities of medical professionals without adding complexity or burden. Given the fragility of MR technology, developing stable, robust, and turnkey solutions that are reliable is a significant challenge that must be addressed to fully realize MR's potential in the OR. I believe that rethinking what XR means in concrete implementations is necessary. For example, while relying on multi-purpose headsets like HL offers the potential for fast, low-effort showcasing, these devices are often over-engineered, but typically leading to fragile solutions. Too often, our thinking is driven by the technological possibilities of these devices, rather than analyzing the needs of surgeons and the underlying processes. I believe specific solutions, especially in terms of hardware, can be much more beneficial by addressing the real needs of surgeons. For instance, using their native tool – the surgical magnification loupe – as an interface for augmented reality, rather than introducing an entirely new device, could provide a more effective solution.

3.14 XR and AI as collaborative tools for shared decision-making

Siriwan Suebnukarn (Thammasat University – Pathum Thani, TH)

XR and AI as collaborative tools for shared decision-making XR enhanced by AI has tremendous potential in shared decision-making (SDM) both preoperative and intraoperative surgical contexts by providing data-driven insights, automating routine tasks, and offering predictive analytics to support clinicians in making informed decisions. Using XR, surgeons can interact with anatomical models in a virtual environment, simulating different surgical approaches, testing possible complications, and adjusting the plan in a shared decision-making context. AI can also suggest the most efficient path based on patient-specific factors. AI can process large datasets from similar cases to predict potential risks (e.g., bleeding, infection, recovery time) for a given patient. During shared decision-making, these risk assessments can be integrated into XR, allowing the surgical team to visualize how different decisions (such as incision placement, resection margins) might impact patient outcomes in real-time. In cases where multiple specialists must collaborate on a complex procedure, XR can help them rehearse the surgery together by simulating potential complications or variations, and specialists can practice responding to these challenges in an immersive XR environment before surgery. Preoperative SDM often involves consultations with patients and families. AI-driven XR systems can present the patient's condition and potential surgical options in an easily understandable format, improving communication and helping the patient become an active participant in the decision-making process. The intraoperative phase is highly dynamic, requiring real-time decision-making based on changing conditions. During surgery, AI can process real-time imaging data (such as live CT, MRI, or ultrasound) and use XR to overlay crucial information onto the surgical field. This allows the surgical team to make shared decisions without losing focus on the procedure. AI can highlight critical structures or suggest optimal surgical paths, reducing the risk of errors. AI can provide real-time decision support by analyzing the evolving intraoperative environment and suggesting changes to the surgical plan. For example, if unexpected bleeding occurs or tissue appears different from preoperative images, AI can process these changes and recommend adjustments, which are then visualized in XR. Surgeons and other specialists can review these suggestions and decide collaboratively on the next steps. By merging the analytical power of AI with the immersive capabilities of XR, teams can enhance collaboration, precision, and patient outcomes in both preoperative planning and intraoperative execution.

3.15 Virtual and Augmented Reality in a Surgical Setting – Research at the University Clinic for Visceral Surgery, Pius-Hospital Oldenburg

Verena Uslar (Universität Oldenburg, DE)

To compensate for the growing personnel shortage in medicine, the possibilities of digitalization must be fully exploited. However, significant problems still exist in transferring new digital tools into clinical practice. Often, medical staff are skeptical about introducing new tools into daily clinical routine and/or the tools do not meet the actual needs.

Therefore, the University Hospital for Visceral Surgery at Pius-Hospital, with its research focus, is dedicated to the user-centered development of digital tools in collaboration with developers and companies from the medical technology industry. This ensures that the tools are developed in a needs-based manner from the outset and increases the likelihood of the tools developed in research projects being transferred to a marketable medical product. At the Dagstuhl Seminar XR4OR, I was pleased to have the opportunity to present three best-practice projects and their results. In the VIVATOP project, we researched VR and AR applications for preoperative planning, intraoperative support, and training and further education with our project partners. For example, a VR planning tool was developed that uses a haptically correct liver model instead of conventional controllers for interaction (see Fig. 1). Additionally, an AR application was researched that allows users to view segmented, patient-specific 3D liver models and can also be used intraoperatively as a navigation aid for tumor resection. The follow-up project AHrEZ is currently working on the automatic segmentation of the liver model based on patients' CT or MRI scans and the automatic placement of the model in the surgical site during the operation.

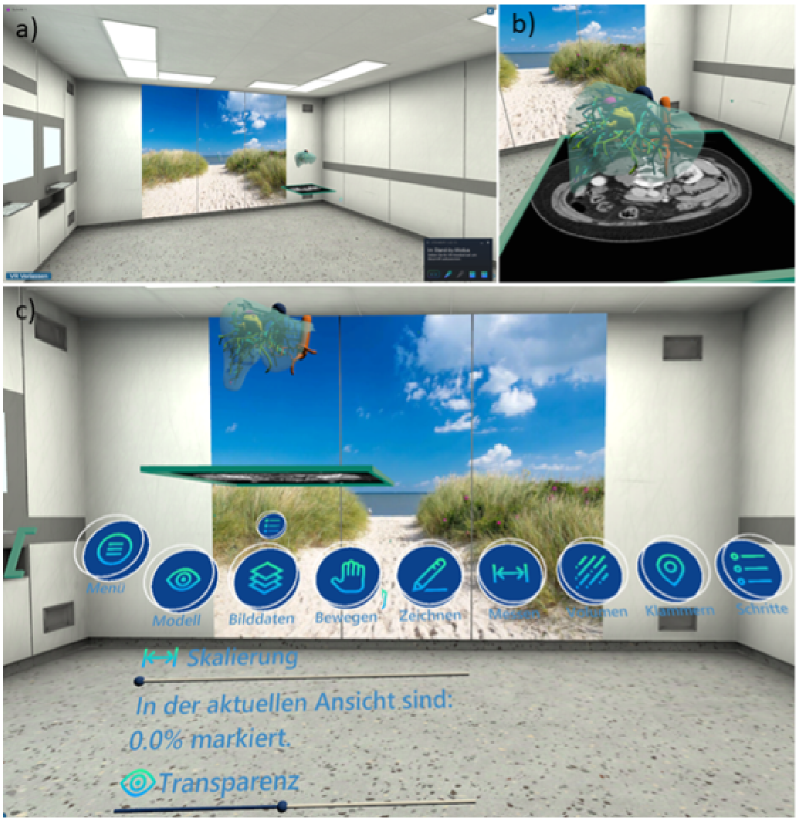

Figure 1 Surgical planning tool (VR) using patient-specific 3D liver models.

For the evaluation of these tools, the University Hospital for Visceral Surgery uses not only the usual subjective measures known from user-centered development, such as questionnaires or interviews, but also new methods like mobile EEG or skin conductance (EDA). Furthermore, the University Hospital operates two Living Labs, where the developed tools can be iteratively tested (see Fig. 2). This allows us to provide our development partners with a comprehensive overview of factors such as feasibility and usability of the developed tools, as well as workload and stress when using the tools, in research projects.

This leads to demonstrators, such as the newly developed lighting concept for the operating room in the SmartOT project (see Fig. 3), being researched in a sustainable manner, so that the implementation as a medical product actually occurs after the project.

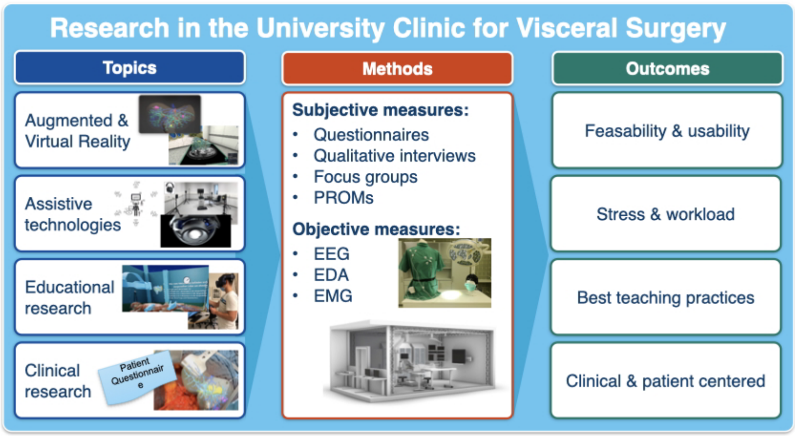

Figure 2 Research at the University Hospital for Visceral Surgery at Pius-Hospital Oldenburg.

Figure 3 Photo of the demonstrator for a new operating room lighting system. The prototype will be installed in one of the new operating rooms at Pius-Hospital this year.

The Dagstuhl Seminar was a confirmation for me that this type of research is urgently needed. Additionally, it had a lasting impact on future research projects through the creation of new contacts.

3.16 Awake Craniotomy

Gregory F. Welch (University of Central Florida – Orlando, US)

The following is an idea I wrote up to share, but then discovered that it is already being done, at least to some extent. Still, I like the idea, and think more could be done in this area.

Awake neurosurgery , also known as awake craniotomy , is a specialized surgical procedure performed while the patient is conscious. This technique is often used when operating on brain regions responsible for critical functions like language, motor control, and memory, to monitor and preserve the patient's cognitive and functional abilities in real-time.

Although I am not a medical professional, I have seen this in person. My mother suffers from Essential Tremor (ET) – a neurological disorder that causes involuntary, rhythmic shaking, typically affecting the hands but sometimes impacting other parts of the body, such as the head, voice, arms, or legs. I was present in the operating room when she had a Deep Brain Stimulation (DBS) system implanted for the first time, and again when she had revision surgery – surgical repositioning of the leads / contacts in her brain.

My thinking is that an awake neurosurgery patient could wear a VR system running an application that automatically, with the guidance of the surgeon, prompts the patient to think and speak about things, perhaps even to carry out cognitive and motor activities. The goal would be to exercise different regions of the brain during the surgery, to increase the chance of detecting any neurological issues caused by the surgery.

For example, much like VR is used for physical therapy, the system could present the patient with nearby virtual targets to look at or even "touch". This could be going on concurrently while the surgeon is operating, normally cycling through some critical activities (cognitive functions), and / or carrying out activities that exercise parts of the brain called out by the surgeon. While the surgeon could call out the specific activities, if they instead call out the goal for the cognitive activity, the system could undertake a broad series of activities that will stimulate or exercise the relevant brain activity.

3.17 Haptics for the OR

Rene Weller (Universität Bremen, DE)

Joint work of Rene Weller, Gabriel Zachmann, Maximilian Kaluschke, Siriwan Suebnukarn, Peter Haddawy
Main reference Maximilian Kaluschke, René Weller, Myat Su Yin, Benedikt W. Hosp, Farin Kulapichitr, Siriwan Suebnukarn, Peter Haddawy, Gabriel Zachmann: "Reflecting on Excellence: VR Simulation for Learning Indirect Vision in Complex Bi-Manual Tasks", in Proc. of the IEEE Conference Virtual Reality and 3D User Interfaces, VR 2024, Orlando, FL, USA, March 16-21, 2024, pp. 712–721, IEEE, 2024.
URL https://doi.org/10.1109/VR58804.2024.00091

Haptic feedback plays a crucial role in realistic surgical simulation and has the potential to enhance actual surgical procedures. We will detail the development of a VR simulator for total hip arthroplasty (THA) and dental surgeries, including the haptic devices, rendering methods, and evaluation of its realism and usability. Furthermore, we discuss how haptic interfaces and devices can be applied in the operating room, focusing on how providing tactile feedback to surgeons can improve precision, safety, and control, ultimately leading to better surgical outcomes.

3.18 Remotely guided robotic surgery through extended reality

Sudanthi Wijewickrema (University of Melbourne, AU)

In locations where expert surgeons are not available, it will be beneficial if remote surgery is possible. This will require the following.

Obtain a model of the region of interest: Pre-operative scans of the patient (regions of interest for the surgery) need to be obtained first. These scans should then be segmented to identify anatomical structures and/or pathologies. Artificial intelligence methods such as deep learning can be used for this. From these segmentations, a 3D model of the region of interest can be created.

Generate a surgical plan: Using the pre-operative 3D model, a surgical plan and optimal path for the surgery should be generated. The path should be displayed on the 3D model in a step-by-step manner to enable real-time guidance.

Register the patient with the pre-operative model: Fiducials can be used as orientation landmarks to register the patient with the 3D model and surgical plan. Registration should be done prior to the start of the surgery and in real-time to ensure the 3D model is aligned with the patient.

Perform remote robotic surgery: The surgeon in the remote location can operate on the 3D model projected via a XR headset and haptics that simulate the robotic surgical system. The operation will be reflected on the actual patient mimicking the actions of the remote surgeon.

4 Open problems

4.1 Dagstuhl Seminar 2025 – XR in the OR: "Pain Points"

Dirk Weyhe (Universität Oldenburg, DE)

Main reference Weyhe D. (2025): Neue digitale Assistenzsysteme: Potenziale in der Viszeralmedizin. Die Chirurgie. 96, 1-2. doi: 10.1007/s00104-024-02220-9
URL 10.1007/s00104-024-02220-9

Introduction. The future of medicine in the Western world is shaped by demographic change and increasingly advanced, individualized, and interdisciplinary therapy options. In contrast, there is a growing shortage of personnel. The Wi-FOR Institute, commissioned by Price Waterhouse Coopers (PWC) Germany, analyzed that in 2022, there was a shortage of 290,000 unfilled positions (6.8%) in the healthcare sector, which is expected to grow to 1.8 million (35.4%) by the year 2035. The Robert Bosch Foundation published a study in 2021 titled "Health Centers for Germany: How a Restart in Primary Care Can Succeed." Their projections indicate that by 2035, rural areas, such as Northwest Germany, will face dramatic undersupply with a primary care coverage rate of less than 75%. The German Medical Association describes in the report on "Resilience in Healthcare" from 2023 by the Expert Council that the increasing demand for medical expertise will be met with a future shortage of doctors.

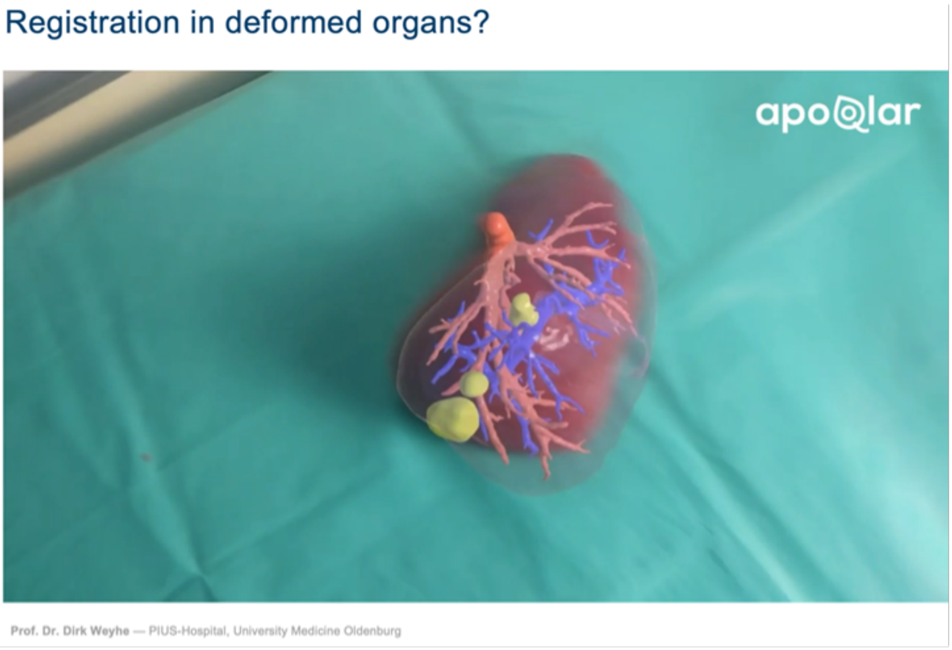

Figure 4 3D printing of a deformable liver with vascular tree projection (yellow – metastases of colorectal carcinoma).

Method. The potentials of digital assistance systems such as AR and VR are presented using the example of the BMBF-funded collaborative project VIVATOP (Weyhe, 2025). In this project, preoperative liver surgeries were planned in VR based on patient-specific segmented DICOM data (CT scans), supported intraoperatively by AR, and the case series was transferred into a database for postoperative debriefing and training purposes. Initial combinations of 3D-printed livers and registered AR projections for training purposes are presented (Fig. 4).

As "pain points" for potential routine application, "Fluid XR in daily use", "automatic placement" (registration), fully automatic segmentation (3D), "haptic feedback" in planning & training systems, and very important digital concepts to compensate for the lack of healthcare workers have been defined.

Summary

The potential of digital assistance systems in visceral surgery is enormous; however, a number of technical solutions are still needed to fully realize these potentials in routine practice with high added value. In this context, telemedicine plays a crucial role and has made significant advancements in recent years.

References

1 Dirk Weyhe. *Neue digitale Assistenzsysteme: Potenziale in der Viszeralmedizin*. Die Chirurgie. 96, 1-2. doi: 10.1007/s00104-024-02220-9

Participants

- Philippe C. Cattin
Universität Basel – Allschwil, CH
- Tanja Döring
Universität Bremen, DE
- Simon Drouin
ETS – Montreal, CA
- Gabor Fichtinger
Queen's University – Kingston, CA
- Tabea Flügge
Charité – Berlin, DE
- Peter Haddawy
Mahidol University – Salaya, TH
- Anja Hennemuth
Charité – Berlin, DE
- Joaquim A. Jorge
University of Lisbon, PT
- Ron Kikinis
Harvard Medical School – Boston, US
- Luv Kohli
InnerOptic Technology – Hillsborough, US
- Mario Lorenz
TU Chemnitz, DE
- Rainer Malaka
Universität Bremen, DE
- Javier Pascau
Carlos III University of Madrid, ES
- Dirk Reiners
University of Central Florida – Orlando, US
- Anke Reinschlüssel
Universität Konstanz, DE
- Andrea Schenk
Fraunhofer MEVIS – Bremen, DE
- Falko Schmid
Heinrich-Heine-Universitätsklinikum Düsseldorf, DE
- Siriwan Suebnukarn
Thammasat University – Pathum Thani, TH
- Verena Uslar
Universität Oldenburg, DE
- Gregory F. Welch
University of Central Florida – Orlando, US
- Rene Weller
Universität Bremen, DE
- Dirk Weyhe
Universität Oldenburg, DE
- Sudanthi Wijewickrema
University of Melbourne, AU
- Gabriel Zachmann
Universität Bremen, DE

Report from Dagstuhl Seminar 25071

Dealing with Complexities in Auction and Matching Market Design

Martin Bichler[*1], **Péter Biró**[*2], **Tom Demeulemeester**[†3], **and Bettina Klaus**[*4]

1 **TU München – Garching, DE.** bichler@in.tum.de
2 **HUN-REN KRTK – Budapest, HU & Corvinus University of Budapest, HU.** biro.peter@krtk.mta.hu
3 **University of Lausanne, CH.** tom.demeulemeester@unil.ch
4 **University of Lausanne, CH.** bettina.klaus@unil.ch

Abstract

Dagstuhl Seminar 25071 gathered an interdisciplinary group of researchers from economics, computer science, and operations research to address current challenges in auction and matching market design. These centralized allocation mechanisms – used in school admissions, kidney exchanges, refugee resettlement, power markets, and spectrum auctions – must increasingly accommodate complex real-world requirements, including multi-objective optimization, dynamic participation, and strategic behavior under uncertainty.

Seminar February 9–14, 2025 – https://www.dagstuhl.de/25071
2012 ACM Subject Classification Theory of computation → Algorithmic game theory and mechanism design
Keywords and phrases algorithms, auctions, game theory, market design, matching markets
Digital Object Identifier 10.4230/DagRep.15.2.41

1 Executive Summary

Martin Bichler (TU München – Garching, DE)
Péter Biró (HUN-REN KRTK – Budapest, HU & Corvinus University of Budapest, HU)
Bettina Klaus (University of Lausanne, CH)

The Dagstuhl Seminar emphasized the intersection of markets with and without monetary transfers, aiming to identify conceptual and methodological tools that can be transferred across auction and matching domains. It provided a platform to exchange insights across disciplines, discuss open problems, and foster new collaborations.

Structure and Activities

The week-long seminar began with participant introductions and featured a carefully balanced mix of:

- Survey talks, which provided foundational overviews and highlighted key developments in the field.

[*] Editor / Organizer
[†] Editorial Assistant / Collector

- Research presentations, focusing on current projects, algorithmic innovations, and real-world applications.
- Rump sessions to stimulate open discussion on challenges and open problems.
- Working group sessions for in-depth exploration of emerging research directions.
- A closing discussion including updates of working groups' results, as well as a social excursion to foster informal interactions.

Highlights of Survey Talks

The seminar featured six in-depth survey talks, scheduled across the week, which framed the discussions and established key interdisciplinary touchpoints:

- Haris Aziz opened with a survey on fair allocation and matching, reviewing fundamental fairness concepts and recent advances in two-sided matching models.
- Thanh Nguyen gave a survey on relaxations in market design, highlighting how continuous relaxations and rounding can be used to tackle complex constraints in matching, exchange, and allocation markets.
- Axel Ockenfels presented a survey on behavioral market design, emphasizing the role of bounded rationality and empirical behavior in designing effective institutions, particularly in the face of global challenges.
- Bettina Klaus delivered a commemorative and forward-looking talk titled "Happy 50(+1)th birthday TTC!", tracing the evolution of the Top Trading Cycles mechanism and its extensions to more complex and realistic preference domains.
- Margarida Carvalho concluded the seminar's survey series with a talk on mathematical optimization for matchings, exploring mixed-integer programming formulations and their role in addressing many-to-one matching problems under constraints.

These survey talks spanned theoretical, algorithmic, and behavioral dimensions of market design and emphasized the importance of cross-disciplinary approaches to current market design problems.

Key Themes and Contributions

Connecting Auction and Matching Domains

Participants explored how auction-based concepts such as artificial pricing, budget constraints, and learning dynamics could be translated into matching settings like course allocation or kidney exchanges. Pseudo-market mechanisms and matching with contracts emerged as promising hybrid models for allocation in settings with soft monetary constraints or policy-based contracts.

Addressing Real-World Complexity

Multiple talks and working groups addressed the need to design mechanisms that account for real-world constraints – ranging from congestion aversion in school choice, to multi-criteria optimization in kidney exchange, to crowd-based last-mile delivery. Algorithmic techniques included mixed-integer programming, approximation algorithms, and simulations tailored to specific institutional settings.

Strategic Behavior and Bounded Rationality

Several presentations focused on modeling and mitigating strategic manipulation in allocation mechanisms. Discussions included satisficing equilibria, perverse incentives created by institutional rankings, and incentive-compatible mechanisms that balance fairness and efficiency. These considerations are essential for ensuring adoption and trust in designed systems.

Emerging Methodologies

Participants presented innovative tools from behavioral economics, agent-based modeling, and game-theoretic learning, alongside traditional mechanism design and algorithmic techniques. Working groups explored algorithmic approaches to topics like financial network compression, weak preference handling in TTC, and equilibrium computation in Bayesian settings.

Conclusion

Dagstuhl Seminar 25071 succeeded in fostering rich interdisciplinary exchange around the growing complexity of modern allocation problems. By emphasizing the connections between auctions and matching markets, and grounding theory in practical application domains, the seminar helped lay the groundwork for future collaboration and innovation in computational market design.

The diversity of topics and expertise reflected both the breadth of the field and the need for integrated approaches to market design challenges in an increasingly complex world.

The organisers thank all the Dagstuhl staff members for their professional support and the participants for enriching the seminar.

2 Table of Contents

Working groups

3 Overview of Talks

3.1 A Survey on Fair Allocation and Matching

Haris Aziz (UNSW – Sydney, AU)

In this survey talk, I discuss some recent results and trends in the literature on fair allocation. I also discuss the connections with two sided matching.

3.2 Learning in Games

Martin Bichler (TU München – Garching, DE)

We analyze the convergence of gradient-based learners to (Bayesian) Nash equilibrium in market games. While equilibrium can be verified ex-post, we discussed ex-ante conditions for convergence in games.

3.3 Multi-criteria optimisation in the Spanish kidney exchange programme

Péter Biró (HUN-REN KRTK – Budapest, HU & Corvinus University of Budapest, HU)

We study the relevance of the order of the objective criteria in the hierarchic optimisation in the Spanish kidney exchange programme. We conduct computer simulations on three real kidney exchange instances from Spain in 2019 and also on simulated datasets by using the ENCKEP kidney exchange simulator. We consider six optimisation criteria that are relevant in the Spanish policy: maximisation of (1) the number of transplants, (2) a composite quality-priority score, (3) the number of cycles, (4) the number of two-cycles and three-cycles with embedded two-cycles, (5) the number of back-arcs in the cycles, and (6) the minimisation of the number of ABO-incompatible transplants. In our simulations we compute all the alternative solutions that can be obtained by a hierarchical optimisation strategy using these criteria, and we analyse the trade-offs with regard to these objectives.

3.4 Sharing with Frictions: Limited Transfers and Costly Inspections

Federico Bobbio (Northwestern University – Evanston, US)

Joint work of Federico Bobbio, Randall Berry, Michael Honig, Thanh Nguyen, Vijay Subramanian, Rakesh Vohra

We study an optimal mechanism design problem for resource sharing between an incumbent noncommercial user and a commercial entrant. Motivated by applications such as wireless spectrum allocation and access to national resources, where noncommercial and commercial users compete for limited availability, the mechanism is subject to two key constraints. First, direct transfers cannot take place to compensate the incumbent for externalities caused by the activity of the commercial user. Instead, the incumbent reports a disutility for sharing to a regulator; the second constraint is that this disutility can only be verified via a costly inspection. The regulator optimizes total welfare by announcing a probability of shared resource assignment along with a probability of inspection. We show that in the optimal mechanism, both allocation and inspection decisions are binary: The regulator either fully allows or prohibits sharing, and inspections occur only for certain reported values when sharing is denied. We further reformulate the problem as a knapsack model to characterize threshold extremal points.

3.5 Mathematical Optimization for Matchings

Margarida Carvalho (University of Montreal, CA)

In this talk, we will begin by reviewing key aspects of effectively solving mixed-integer programming (MIP) formulations in practice. We will then explore optimization problems in many-to-one matchings, gradually progressing from computationally tractable cases to NP-hard ones. Along the way, we will examine MIP models, emphasizing the interplay between modeling techniques and solution strategies. Finally, we will conclude with open questions on the role of mathematical optimization in matching markets.

3.6 Bidding for subsidies with one's patience

Gian Caspari (Zentrum für Europäische Wirtschaftsforschung – Mannheim, DE)

Main reference Gian Caspari: "Bidding for Subsidies with One's Patience". Available at SSRN
URL https://dx.doi.org/10.2139/ssrn.5133595

We study the problem of distributing subsidies in a market that includes both marginal individuals in need of assistance and infra-marginal individuals who would purchase the subsidized product without additional incentives. We propose the use of a wait time auction, where individuals bid the amount of time they are willing to wait in exchange for a specified subsidy amount. This design enables more direct targeting of marginal individuals, thereby enhancing the overall effectiveness of the subsidy program. Furthermore, screening is costless in equilibrium as no wait times are imposed, and practical robustness against deviations from equilibrium behavior can be ensured by implementing a maximum allowable bid.

3.7 Fairness in Assignments with Congestion-Averse Agents: Concepts, Algorithms, and Complexity

Jiehua Chen (TU Wien, AT)

The congested assignment problem is concerned with assigning agents to posts where agents care about both the posts and their congestion levels. Here, agents are averse to congestion, consistently preferring lower over higher congestion for the same resource. Such scenarios are prevalent across many domains, including traffic management and school choice, where fair resource allocation is crucial. Congested assignment can be considered as a restricted variant of the Group Activity Selection problem, introduced by Darmann et al. Additionally, it is related to many-to-one matching in matching under preferences.

In this talk, I will explore one ex-ante fairness concept, top-fairness, and two ex-post fairness concepts, envy-freeness and competitiveness. The top-fairness and competitiveness concepts were recently introduced by Bogomolnaia and Moulin. While a top-fair or envy-free assignment always exists and can be found easily, competitive assignments do not always exist. The talk will cover the following key points:

1. An efficient method to determine the existence of competitive or maximally competitive assignments for a given congestion profile.
2. Two optimization variants of congested assignments and their computational complexity: a) Finding a top-fair assignment that is envy-free b) Finding a top-fair assignment that is maximally competitive. Both variants are NP-hard, unfortunately.
3. Parameterized algorithms for these NP-hard problems.

3.8 Welfare Lower Bounds in House Allocation Problems with Existing Tenants: A Characterization

Yang Chen (University of Lausanne, CH)

Joint work of Yang Chen, Di Feng
Main reference Chen, Yang and Di Feng: "Welfare Lower Bounds in House Allocation Problems with Existing Tenants: A Characterization". Available at SSRN, 2025.
URL https://dx.doi.org/10.2139/ssrn.5111246

We study Abdulkadiroğlu and Sönmez's (1999) house allocation problems with existing tenants. In particular, we focus on three welfare lower bounds conditions for agents, with respect to tenants' endowments, solidarity for newcomers, and fairness for all agents. Based on these three welfare lower bounds conditions, together with three other well studied properties, namely pair efficiency, strategy-proofness and weak neutrality, we characterize a class of mechanisms proposed by Sönmez and Ünver (2005): the You Request My House-I Get Your Turn mechanism with Newcomer Priorities (YTP), which is a hybrid between the top trading cycles mechanism and a serial dictatorship mechanism of newcomers. Our result constitutes the first characterization, in the absence of invariance properties, e.g., non-bossiness and consistency.

3.9 Complexity and Manipulation of International Kidney Exchange Programs with Country-Specific Parameters

Rachael Colley (University of Glasgow, GB)

Joint work of Rachael Colley, David Manlove, Daniël Paulusma, Michelle Zhang

Kidney Exchange Programs (KEPs) facilitate the exchange of kidneys, and larger pools of recipient-donor pairs tend to yield proportionally more transplants, leading to the proposal of international KEPs (IKEPs). However, practical limitations must be considered in IKEPs to ensure that countries remain willing to participate. Thus, we study IKEPs with country-specific parameters, restricting the selected transplants to be feasible for the countries to conduct, e.g., imposing an upper limit on the number of consecutive exchanges within a country's borders. We provide a complexity dichotomy of finding allocations for different country-specific parameters and study the potential for countries to misreport their parameters to increase their allocation. As manipulation can harm the total number of transplants, we propose an *individually rational* (IR) and *incentive compatible* (IC) mechanism. We first give a theoretical approximation ratio for our mechanism in terms of the number of transplants and then use simulations to show that the cost of IC and IR in practice is between 18% and 34% of the maximum number of transplants.

3.10 Trading homogeneous goods at multiple delivery points

Dávid Csercsik (HUN-REN KRTK – Budapest, HU)

The EU Energy Platform was established in April 2022. As part of the EU Energy Platform the EU also set up a demand aggregation and joint purchasing mechanism (DA/JPM), called AggregateEU. The AggregateEU is a non-binding mechanism that aims to match the supply and demand of natural gas in the context of multiple delivery points (mostly LNG terminals). The non-binding nature of the platform raises several potential problems (like the phenomenon of overbidding and its implications). Still, the synthesis of a non-binding alternative is also non-trivial. We aim to create an auction framework in which participants, either buyers or sellers of a homogeneous good are characterized by limited access and different access costs to each delivery point, while also considering the principle of fairness, i.e. equal access to cheap sources for buyer participants.

3.11 Open Questions from Kidney Exchange and Stable matches

Gergely Csáji (Eötvös Lorand University – Budapest, HU)

In this talk I will give an overview of some recent projects, focusing on the remaining open questions. The topics include NTU games for kidney exchange, core of many-to-many matching markets, student and resident allocation and a new notion called k-stability.

3.12 Pricing Valid Cuts for Price-Match Equilibria

Robert Day (University of Connecticut, US) and Ben Lubin (Boston University, US)

We use valid inequalities (cuts) of the binary integer program for winner determination in a combinatorial auction (CA) as "artificial items" that can be interpreted intuitively and priced to generate Artificial Walrasian Equilibria. While the lack of an integer programming gap is sufficient to guarantee a Walrasian equilibrium, we show that it does not guarantee a "price-match equilibrium" (PME), a refinement that we introduce, in which prices are justified by an iso-revenue outcome for any hypothetical removal of a single bidder. We prove the existence of PME for any CA and characterize their economic properties and computation. We implement minimally artificial PME rules and compare them with other prominent CA payment rules in the literature.

3.13 Fair integer programming under dichotomous and cardinal preferences

Tom Demeulemeester (University of Lausanne, CH)

Joint work of Tom Demeulemeester, Dries R. Goossens, Ben Hermans, Roel Leus
Main reference Tom Demeulemeester, Dries R. Goossens, Ben Hermans, Roel Leus: "Fair integer programming under dichotomous and cardinal preferences", Eur. J. Oper. Res., Vol. 320(3), pp. 465–478, 2025.
URL https://doi.org/10.1016/J.EJOR.2024.08.023

One cannot make truly fair decisions using integer linear programs unless one controls the selection probabilities of the (possibly many) optimal solutions. For this purpose, we propose a unified framework when binary decision variables represent agents with dichotomous preferences, who only care about whether they are selected in the final solution. We develop several general-purpose algorithms to fairly select optimal solutions, for example, by maximizing the Nash product or the minimum selection probability, or by using a random ordering of the agents as a selection criterion (Random Serial Dictatorship). We also discuss in detail how to extend the proposed methods when agents have cardinal preferences. Lastly, we evaluate the proposed methods on the specific application of kidney exchange. We find that while the methods maximizing the Nash product or the minimum selection probability outperform the other methods on the evaluated welfare criteria, methods such as Random Serial Dictatorship perform reasonably well in computation times that are similar to those of finding a single optimal solution.

3.14 Complexities in the Roommates Problem

Frederik Glitzner (University of Glasgow, GB)

In the Stable Roommates problem, the aim is, given a set of agents and their non-bipartite ordinal preferences over each other are given, to compute a matching of the agents such that no two agents prefer each other to their assigned partners (or are unmatched). In this talk, we will review concepts related to this problem, outline new theoretical and empirical perspectives on how to deal with instances that do not admit such matchings, and pose a range of algorithmic, structural, and complexity-theoretic questions.

3.15 (Our) Open practical and theoretical problems in matching

Claus-Jochen Haake (Universität Paderborn, DE)

We discuss two open problems in matching. First, in terms of a farsightedness, is a maximal matching indeed maximal? Second, under which conditions can we guarantee the existence of stable matchings, when preferences are generated through the satisfaction of objective criteria? Here we draw a connection to bargaining theory.

3.16 Stable matchings and distributive lattices

Zsuzsanna Jankó (Corvinus University of Budapest, HU & HUN-REN KRTK – Budapest, HU)

Joint work of Tamás Fleiner, Zsuzsanna Jankó, Babak Miraftab, Konstantinos Stavropoulos

It is a classic result that the set of one-to-one stable matchings form a distributive lattice. We generalize this to many-to-many matchings with substitutable, IRC and increasing choice functions, and show that the lattice of the solutions is distributive. We know one-to-one stable matchings have rotations, we try to look at how to find something similar in many-to-many matchings.

3.17 Housing Market with Identical Objects and Top Trading Cycles Mechanisms

Mehmet Karakaya (Izmir Katip Çelebi University, TR)

Joint work of Mehmet Karakaya, Bettina Klaus, Jan Christoph Schlegel

We consider the housing market model with identical objects where each agent is indifferent among identical objects. The Top Trading Cycles (TTC) mechanism with uniform tie-breaking satisfies individual rationality, Pareto optimality, and strategy-proofness. We introduce a new consistency notion for housing market with identical objects model. We show that when there are two types of objects TTC mechanism with uniform tie-breaking is consistent. However, when there are at least three types of objects TTC mechanism with uniform tie-breaking violates consistency. Our conjecture is that there is no rule satisfying individual rationality, Pareto optimality, strategy-proofness, and consistency for housing market with identical objects.

3.18 Happy 50 (+1) th birthday TTC! A Survey

Bettina Klaus (University of Lausanne, CH)

I present a survey talk commemorating more that 50 years of Gale's famous top trading cycles (TTC) algorithm. The classical housing market model and the TTC algorithm were introduced by Shapley and Scarf (1974) in their seminal paper "Cores and Indivisibilities," Journal of Mathematical Economics 1. The TTC agorithm / rule satisfies many desirable properties relating to weak / strong core stability, efficiency, and incentives, and, remarkably, is characterized by various combinations of these properties. I review some of the most important papers in the field and include some recent results on variations of the Shapley Scarf housing market model (multiple object (re)allocation, limited externalities, weak preference domains).

3.19 Correlation of rankings in matching markets

Patrick Loiseau (Inria – Palaiseau, FR)

Joint work of Rémi Castera, Patrick Loiseau, Bary Pradelski
Main reference Rémi Castera, Patrick Loiseau, Bary Pradelski: "Correlation of Rankings in Matching Markets", 2025.
URL https://hal.science/hal-03672270

We study the role of correlation in matching, where multiple decision-makers simultaneously face selection problems from the same pool of candidates. We propose a model in which decision-makers have varying information on candidates from different sociodemographic groups when evaluating and ranking them, thus leading to varying correlations among candidates' priority scores. Such differential correlation arises, for example, when the cost of information acquisition, decision-maker preferences, or the prevalence of selection criteria vary across sociodemographic groups. We show that a lower correlation for one of the groups worsens the outcome for all groups, thus leading to efficiency loss. Moreover, students from a given group are worse of as their own correlation level increases. This implies that it is advantageous to belong to a low-correlation group. Finally, we extend the extent tie-breaking literature to multiple priority classes and intermediate levels of correlation. Overall, our results point to a previously overlooked systemic source of group inequalities in school, university, and job admissions.

3.20 Perverse incentives created by rankings

Thayer Morrill (North Carolina State University – Raleigh, US)

In many applications, participants care about an exogenous ranking that potentially distorts the preferences they submit in a centralized match. For example, a ranking influenced by a schools yield percentage might lead to a school rejecting students who the school expects to reject them. In the doctor-hospital match, the rank-to-fill metric (what spot on your preference list where you matched to) leads some hospitals to pressure students to rank it first. Can we design rankings which capture the relative quality of the agents being match but which do not distort the agent's preference reports?

3.21 Relaxations in Market Design

Thanh Nguyen (Purdue University – West Lafayette, US)

This talk presents a survey of a new approach to market design for complex environments. The approach first relaxes the discreteness of market design problems into a continuous market equilibrium solution and then rounds it to a nearby solution. I illustrate this method in four different settings: stable matching with complementarities, competitive equilibrium in allocations beyond substitutes and with budget constraints, combinatorial assignment and exchange without monetary transfer, and social choice.

3.22 Behavioral Market Design

Axel Ockenfels (Universität Köln, DE & MPI – Bonn, DE)

Addressing pressing economic and social challenges – such as pandemics and other health crises, climate change, and energy scarcity – requires changes in behavior. In this talk, I will use case studies, primarily from my own research, to illustrate how human behavior and bounded rationality influences the design of institutions aimed at aligning incentives and actions with overarching goals. I will argue that economic design research and behavioral science are often complementary, rather than substitutes, in promoting effective behavioral change.

3.23 Online Bipartite Matching with Replacements

Katarzyna Paluch (University of Wroclaw, PL)

In the online bipartite matching problem with replacements, all the vertices on one side of the bipartition (servers) are given, and the vertices on the other side (clients) arrive one by one with all their incident edges. The goal is to maintain a maximum matching while minimizing the number of changes (replacements) to the matching.

3.24 Satisficing equilibrium

Bary Pradelski (CNRS – Oxford, GB)

Joint work of Bary S. R. Pradelski, Bassel Tarbush
Main reference Bary S. R. Pradelski, Bassel Tarbush: "Satisficing Equilibrium", CoRR, Vol. abs/2409.00832, 2024.
URL https://doi.org/10.48550/ARXIV.2409.00832

We propose a solution concept in which each agent i does not necessarily optimize but selects one of their top k_i actions. Our concept accounts for heterogeneous agents' bounded rationality. We show that there exist satisficing equilibria in which all but one agent best-respond and the remaining agent plays at least a second-best action in asymptotically almost all games. Additionally, we define a class of approximate potential games in which satisficing equilibria are guaranteed to exist. Turning to foundations, we characterize satisficing equilibrium via decision theoretic axioms and we show that a simple dynamic converges to satisficing equilibria in almost all large games. Finally, we apply the satisficing lens to two classic games from the literature.

3.25 Jumping to Conclusions

Sven Rady (Universität Bonn, DE)

This talk presents a continuous-time game of strategic experimentation with two-armed bandits in which payoffs on the risky arm are generated by a compound Poisson process. The distribution of payoff increments depends on an unkown state of the world, whereas the arrival rate of such increments does not. As a consequence, players learn exclusively from the size of the lump-sum payoffs they obtain. Under perfect monitoring, the piecewise constancy of beliefs allows for an analysis of stationary Markov equilibria with essentially discrete-time techniques. In turn, these Markov equilibria constitute perfect Bayesian equilibria when players observe all payoffs, but not each other's actions. Strongly symmetric equilibria cannot improve on the symmetric Markovian one. Whether asymmetric equilibria can achieve higher payoffs than Markovian equilibria is the focus of ongoing work.

3.26 Strategic Queues with Priority Classes

Marco Scarsini (LUISS University – Rome, IT)

Joint work of Maurizio D'Andrea, Marco Scarsini
Main reference Maurizio D'Andrea, Marco Scarsini: "Strategic Queues with Priority Classes", CoRR, Vol. abs/2502.05906, 2025.
URL https://doi.org/10.48550/ARXIV.2502.05906

We consider a strategic M/M/1 queueing model under a first-come first-served regime, where customers are split into two classes and class A has priority over class B. Customers can decide whether to join the queue or balk, and, in case they have joined the queue, whether and when to renege. We study the equilibrium strategies and compare the equilibrium outcome and the social optimum in the two cases where the social optimum is or is not constrained by priority.

3.27 The Strong Core of Housing Markets with Partial Order Preferences

Ildikó Schlotter (HUN-REN KRTK – Budapest, HU)

Joint work of Ildikó Schlotter, Lydia Mirabel Mendoza Cadena
Main reference Ildikó Schlotter, Lydia Mirabel Mendoza Cadena: "The Strong Core of Housing Markets with Partial Order Preferences", CoRR, Vol. abs/2501.15834, 2025.
URL https://doi.org/10.48550/ARXIV.2501.15834

We study the strong core of housing markets when agents' preferences over houses are expressed as partial orders. We provide a structural characterization of the strong core, and propose an efficient and weakly group-strategyproof algorithm that finds an allocation in the strong core or decides that it is empty, even in the presence of forced and forbidden arcs. Additionally, we show that certain results known for the strong core in the case when agents' preferences are weak orders can be extended to the setting with partial order preferences; among others, we show that the strong core in such housing markets satisfies the property of respecting improvements.

3.28 Service Assignment at the United States Naval Academy

Naomi Utgoff (United States Naval Academy – Annapolis, US)

Joint work of Ashwin Kambhampati, Chad Redmer, Naomi Utgoff

This paper studies service assignment at the United States Naval Academy (USNA). Service assignment is the process which assigns each member of USNA's senior class to one of several communities, each representing a career path within the Navy. This matching problem is unique to the literature in that it features hard minimums, community individual rationality constraints, and the requirement that every midshipman be assigned to a community. A stable assignment is generally impossible under these requirements: we characterize a natural relaxation that we call minimal unfairness. We then introduce the individual-rationality adjusted deferred acceptance algorithm (IRDA) and show that it always selects a minimally unfair assignment which respects the Navy's requirements. This mechanism relies on the construction of caps based on the primitives, namely the minimums, maximums, and individual rationality constraints. These caps thus determine when a community must reject a midshipman, as well as which midshipman should be rejected. Unlike standard deferred acceptance in which the rejecting community simply rejects its least favorite proposal(s) above its ceiling, IRDA identifies a rejectable set of midshipmen, and forces the rejection of a midshipman in that set. We also discuss two alternative mechanisms, generalized serial dictatorship (GSD) and eligibility type capped deferred acceptance (ECDA), and contrast their properties with IRDA.

3.29 Lastmile delivery with crowdsourcing

Ana Viana (INESC TEC – Porto, PT)

Last-mile delivery refers to the final leg in the distribution process, typically involving the transportation of goods from a depot or store to the customer's house. Traditionally, this role was fulfilled by professional delivery fleets. However, the exponential growth of e-commerce globally has introduced significant logistical challenges, requesting for innovative delivery models. One such model is based on the concept of crowdsourcing, where individuals – commonly referred to as occasional drivers (ODs) – who have no professional affiliation with the retailers, sign up to deliver goods to customers for a compensation. That is the case of e.g. Amazon Flex.

Several research challenges can be considered. To name a few:

1. How should parcels be assigned for delivery – should they be handled by the professional fleet or an OD? And among ODs, which driver is the best candidate for a specific delivery?
2. Beyond economic efficiency, how can we ensure ODs remain engaged and satisfied with the program
3. What is the optimal payment structure for an OD to ensure fairness while maintaining cost efficiency for the retailer?

The purpose of this talk is to provide a brief introduction to the problem and discuss with the audience potential areas of research aligned with the main topics of the seminar.

3.30 Computing Balanced Solutions for Large International Kidney Exchange Schemes When Cycle Length Is Unbounded

Xin Ye (Durham University, GB)

Joint work of Márton Benedek, Péter Biró, Gergely Csáji, Matthew Johnson, Daniël Paulusma, Xin Ye
Main reference Márton Benedek, Péter Biró, Gergely Csáji, Matthew Johnson, Daniël Paulusma, Xin Ye: "Computing Balanced Solutions for Large International Kidney Exchange Schemes when Cycle Length is Unbounded", in Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2024, Auckland, New Zealand, May 6-10, 2024, pp. 2153–2155, International Foundation for Autonomous Agents and Multiagent Systems / ACM, 2024.
URL https://doi.org/10.5555/3635637.3663091

In kidney exchange programmes (KEP) patients may swap their incompatible donors leading to cycles of kidney transplants. Recently, countries try to merge their national patient-donor pools leading to international KEPs (IKEPs). As shown in the literature, long-term stability of an IKEP can be achieved through a credit-based system. In each round, every country is prescribed a "fair" initial allocation of kidney transplants. The initial allocation, which we obtain by using solution concepts from cooperative game theory, is adjusted by incorporating credits from the previous round, yielding the target allocation. The goal is to find, in each round, an optimal solution that closely approximates this target allocation. There is a known polynomial-time algorithm for finding an optimal solution that lexicographically minimizes the country deviations from the target allocation if only 2-cycles (matchings) are permitted. In practice, kidney swaps along longer cycles may be performed. However, the problem of computing optimal solutions for maximum cycle length ℓ is NP-hard for every $\ell \geq 3$. By contrast, the problem is polynomial time solvable once we allow unbounded cycle length. However, in contrast to the case where $\ell = 2$, we show that for $\ell = \infty$, lexicographical minimization is only polynomial-time solvable under additional conditions (assuming $P \neq NP$). Nevertheless, the fact that the optimal solutions themselves can be computed in polynomial time if $\ell = \infty$ still enables us to perform a large scale experimental study for showing how stability and total social welfare are affected when we set $\ell = \infty$ instead of $\ell = 2$.

4 Working groups

4.1 On the Existence of Strict Bayes-Nash Equilibrium in Bayesian All-Pay Auctions

Martin Bichler (TU München – Garching, DE), Janik Bürgermeister (TU München, DE), Julius Durmann (TU München, DE), Bary Pradelski (CNRS – Oxford, GB), Sven Rady (Universität Bonn, DE), and Marco Scarsini (LUISS University – Rome, IT)

We are interested in the question of why gradient-based no-regret learning converges to the Bayes-Nash equilibrium in Bayesian market games such as auctions, contests, and oligopoly pricing. For this, we first focus on the all-pay auction. The complete-information all-pay auction only has a mixed Nash equilibrium. In contrast, with continuous types and actions, there is a unique pure Bayes-Nash equilibrium (PBNE). After discretizing the type and action space of this continuous game, we usually observe convergence to a PBNE. However, existence and convergence depend on the discretization of the type and action space. We aim to explain the reasons why the complete-information all-pay auction has no PBNE, but the discretized version of the game does. These questions should shed light on the more general reasons why we see convergence to a PBNE in a wide variety of Bayesian games.

4.2 Portfolio compression in financial networks

Péter Biró (HUN-REN KRTK – Budapest, HU & Corvinus University of Budapest, HU)

In a financial network the banks (or companies) have initial assets and liabilities towards each other. Financial clearing means that all the banks have to pay their obligations immediately. If a bank has more dept than its assets and credits then it will go bankrupt, and can only pay part of its liabilities to its creditors using a pre-defined payment rule, potentially causing a cascading effect of bankruptcy in the network involving banks that looked healthy initially. The so-called systemic risk in a network can be even higher if certain bankruptcy costs occur when a bank defaults, generating amplified losses in the economy.

Portfolio compression is an act in financial markets where the financial obligations are cleared in a cycle after the mutual agreement of the agents involved. The coordination of the portfolio compression can be done by private agencies, e.g., TriOptima for banks, or by governmental agencies, as in Romania for companies. When the agents give their approval in advance to a central coordinator for portfolio compression, then the coordinator can choose a portfolio compression over one or multiple cycles. The goal of the coordinator can be to reduce as much dept as possible, or to reduce the systemic risk. In our working group, we discussed the possibility of using a mixed integer linear programming approach with bi-level optimisation to finding a portfolio compression for a given financial network and payment rules to maximise these objectives.

4.3 Potential applications of mechanism design approaches for Last Mile Delivery Problems

Dávid Csercsik (HUN-REN KRTK – Budapest, HU)

In supply chain management, the expression last mile describes the logistical challenges corresponding to the final phase of transportation, in which items are delivered to their terminal destinations. Last-mile delivery (LMD) is the most expensive and most polluting part of the supply chain. Considering the totality of simultaneous LMDs, efficiency is not ideal. Multiple vehicles take part in the process, partially loaded. It is however not trivial, how the coordination of LMD operations may be increased, without monopolizing the sector.

The concept of cooperation in the delivery process is not novel. Literature sources emphasize the importance of central planning (or collaborative planning), which reduce the the overall cost and so the sum of distance and increase the efficiency at the same time. However, in these works the collaboration platform is operated by delivery companies and focuses mainly on profit sharing mechanism, or in other word the players in the game are the the delivery companies and the single delivery tasks are not considered as active members of the market.

As a particular approach, auctions may be efficiently used to increase the coordination in transportation management, more specifically in the efficient procurement of transportation services. The first optimization-based auction for a transportation service allocation, which allowed package (or combined) bids dates to 1992. The items to be auctioned were service rights for lanes, i.e., single transportation paths between cities, and the service rights have been allocated for years.

4.4 TTC rules for weak preferences

Bettina Klaus (University of Lausanne, CH)

For Shapley-Scarf housing markets with weak preferences the TTC rule with fixed tie-breaking is not Pareto efficient. Ehlers (2014): "Top trading with fixed tie-breaking in markets with indivisible goods," Journal of Economic Theory, shows that Gale's top trading cycles algorithm with fixed tie-breaking is characterized by weak efficiency, individual rationality, strategyproofness, non-bossiness, and consistency.

For Shapley-Scarf housing markets with strict preferences, the TTC rule is the only rule satisfying Pareto efficiency, individual rationality, and strategy-proofness. A natural question then is: Are there generalizations of TTC that are satisfying the Ma properties? The answer is yes, as illustrated by the following references: Jaramillo and Manjunath (2012); Alcalde-Unzu and Molis (2013); Aziz and de Keijzer (2012); Plaxton (2013); Saban and Sethuraman (2013); Xiong, Wang, and He (2022). Since there are several extensions of TTC to the weak preference domain, can we refine the class by imposing more properties?

Our working group (Bettina, Haris, Mehmet, Naomi, Tamas, Thayer, Yang) first studied the Xiong, Wang, and He (2022) algorithm. We conjecture that it satisfies strong core stability (proof to be worked out) and plan to analyze the other properties for it (and for the other mechanisms). We also brainstormed on the general structure behind all proposed algorithms and their link to the underlying properties. We'll continue with these ideas and hopefully will obtain some neat results.

4.5 Gaming the DA in Amsterdam

Thilo Klein (Zentrum für Europäische Wirtschaftsforschung – Mannheim, DE)

School authorities increasingly turn to lotteries as a means of addressing school segregation. In Europe, a growing number of cities have adopted lottery-based admissions, replacing traditional priority-based systems.

Our working group (Federico, Margarida, Gian, Jiehua, Gergely, Tom, Frederik, Claus-Jochen, Zsuzsanna, Thilo, Katarzyna, Nadja) examined the case of Amsterdam, which introduced the Deferred Acceptance (DA) mechanism with multiple tie-breaking and no priorities in 2015. We analysed the city's responses to the challenges that emerged from this approach and identify three key areas for improvement: (1) Efficiency: In response to legal disputes from parents seeking to swap school places, Amsterdam transitioned to a single tie-breaking system. (2) Inequality: To address concerns about unequal access, the city introduced a guarantee of placement at a top-k school from students' ranked preferences. (3) Reciprocity: Amsterdam added a priority rule for students ranking a school first, effectively introducing an Immediate Acceptance (IA) step before the DA process. While these modifications, implemented in 2016 and 2018, aimed to improve outcomes, they rendered the mechanism obviously manipulable.

We analysed an adaptive mechanism combining a maximum b-matching with the rank minimizing (RM) mechanism. This combination balances efficiency, equality, and reciprocity while muting strategic incentives. We will continue this analysis and discuss it with researchers and policy makers in Amsterdam.

4.6 Minimum spanning subgraph for a group of agents

Dušan Knop (Czech Technical University – Prague, CZ)

Consider an undirected graph $G = (V, E)$ and a set of agents $\mathcal{N}$.Each agent $i \in \mathcal{N}$ has a valuation function over the graph's edges $w_i : E \to \mathbb{R}$. Our problem is to find a spanning subgraph $H = (V, F)$ of G such that $|F|$ is minimal and for each agent $i \in \mathcal{N}$, the graph H contains a minimal spanning tree of G with respect to w_i.

We started with the case where $|\mathcal{N}| = 2$, and provided two polynomial-time algorithms for the problem. The first algorithm relies on the known fact that the edge sets of minimum spanning trees for some valuation function w_i form a matroid, and uses a weighted matroid-intersection algorithm to compute in polynomial time a largest edge set that is contained in a w_i-minimum spanning tree for both $i = 1, 2$. Such an edge set can then be extended to a solution of minimum size in a straightforward manner. The second algorithm is combinatorial and relies only on elementary techniques and a few structural observations without building on matroid theory.

4.7 GenAI for Bid Language and Bid Entry Design

Ben Lubin (Boston University, US), Robert Day (University of Connecticut, US), and Thanh Nguyen (Purdue University – West Lafayette, US)

The initial steps and focus areas of a project aimed at integrating generative AI into auction bidding systems. After exploring multiple potential use cases, the team chose to prioritize building a bidding interface for a combinatorial auction, deferring broader mechanism design-related applications to future work. The team explored applications such as TV advertising slot auctions and school choice mechanisms, and reviewed existing bidding language models including combinatorial assignments with budget constraints. The plan is to begin with simple sealed bid scenarios using value queries, avoiding more complex iterative approaches initially.

Preliminary testing with various versions of Claude and GPT models revealed limitations and strengths. While the models show potential in templating and general language understanding, they struggle with advanced quantitative reasoning. GPT-4's reasoning model was marginally better but still insufficient for more nuanced tasks. The team observed a bias toward online advertising tasks – likely due to the models' training data – further highlighting the need for customization to suit a specific bidding application (actually to include advertising, as the model biases towards contemporary mechanisms like adsense).

To address these challenges, the team recognized a need for customization, likely both in the LLM and in the code around it. More specifically, this likely includes integrating domain-specific information via Retrieval-Augmented Generation (RAG), building a custom value calculator (potentially with LLM support), and creating agentic systems that combine LLMs with tool-use capabilities. Fine-tuning the LLM is a potential step, though we are hoping to avoid having to do this.

The initial work at the conference was promising, but also made clear that there is quite a bit of follow on work necessary to get a usable system. We also briefly discussed how such a system might be evaluated as a path to a publishable paper.

Participants

- Haris Aziz
UNSW – Sydney, AU
- Martin Bichler
TU München – Garching, DE
- Péter Biró
HUN-REN KRTK – Budapest, HU & Corvinus University of Budapest, HU
- Federico Bobbio
Northwestern University – Evanston, US
- Janik Bürgermeister
TU München, DE
- Margarida Carvalho
University of Montreal, CA
- Gian Caspari
Zentrum für Europäische Wirtschaftsforschung – Mannheim, DE
- Jiehua Chen
TU Wien, AT
- Yang Chen
University of Lausanne, CH
- Rachael Colley
University of Glasgow, GB
- Gergely Csáji
Eötvös Lorand University – Budapest, HU
- Dávid Csercsik
HUN-REN KRTK – Budapest, HU
- Robert Day
University of Connecticut, US
- Tom Demeulemeester
University of Lausanne, CH
- Julius Durmann
TU München, DE
- Tamás Fleiner
Budapest University of Technology & Economics, HU
- Frederik Glitzner
University of Glasgow, GB
- Claus-Jochen Haake
Universität Paderborn, DE
- Zsuzsanna Jankó
Corvinus University – Budapest, HU
- Mehmet Karakaya
Izmir Katip Çelebi University, TR
- Bettina Klaus
University of Lausanne, CH
- Thilo Klein
Zentrum für Europäische Wirtschaftsforschung – Mannheim, DE
- Xenia Klimentova
INESC TEC – Porto, PT
- Dušan Knop
Czech Technical University – Prague, CZ
- Patrick Loiseau
Inria – Palaiseau, FR
- Ben Lubin
Boston University, US
- Thayer Morrill
North Carolina State University – Raleigh, US
- Thanh Nguyen
Purdue University – West Lafayette, US
- Axel Ockenfels
Universität Köln, DE & MPI – Bonn, DE
- Katarzyna Paluch
University of Wroclaw, PL
- Bary Pradelski
CNRS – Oxford, GB
- Sven Rady
Universität Bonn, DE
- Baharak Rastegari
University of Southampton, GB
- Marco Scarsini
LUISS University – Rome, IT
- Ildikó Schlotter
HUN-REN KRTK – Budapest, HU
- Nadja Stroh-Maraun
Universität Paderborn, DE
- Naomi Utgoff
United States Naval Academy – Annapolis, US
- Carmine Ventre
King's College London, GB
- Ana Viana
INESC TEC – Porto, PT
- Xin Ye
Durham University, GB

Report from Dagstuhl Seminar 25072

What You Hear is What You See? Integrating Sonification and Visualization

Wolfgang Aigner*[1], Sara Lenzi*[2], Niklas Rönnberg*[3], Kajetan Enge†[4], and Alexander Rind†[5]

1 **St. Pölten University of Applied Sciences, AT.** wolfgang.aigner@fhstp.ac.at
2 **University of Deusto – Bilbao, ES; Ikerbasque Basque Foundation for Science – Bilbao, ES.** sara.lenzi@deusto.es
3 **Linköping University, SE.** niklas.ronnberg@liu.se
4 **Universität für Musik & darstellende Kunst – Graz, AT.** kajetan@enge.at
5 **FH – St. Pölten, AT.** Alexander.Rind@fhstp.ac.at

Abstract

This report summarizes the outcomes of a Dagstuhl Seminar that brought together international experts from the data visualization and data sonification communities. The seminar aimed to bridge these two fields by exploring their complementary potential in enhancing human-data interactions. Over five days, participants engaged in discussions, inspiration sessions, and presentations focused on nine key topics, including design theory, application examples, perception, interaction, and evaluation of integrated visualization and sonification. The seminar facilitated a collaborative environment in which both communities shared knowledge, identified common challenges, and began defining a novel field that combines visual and auditory data representation. The event marked a significant step towards establishing a common ground between these disciplines, with the goal of improving data analysis and presentation in an increasingly data-driven world.

Seminar February 9–14, 2025 – https://www.dagstuhl.de/25072
2012 ACM Subject Classification Human-centered computing → Visualization; Human-centered computing → Auditory feedback; Human-centered computing → Sound-based input / output; Applied computing → Sound and music computing
Digital Object Identifier 10.4230/DagRep.15.2.63

1 Executive Summary

Wolfgang Aigner (St. Pölten University of Applied Sciences, AT)
Sara Lenzi (University of Deusto – Bilbao, ES & Ikerbasque Basque Foundation for Science – Bilbao, ES)
Niklas Rönnberg (Linköping University, SE)

In our daily lives, we as humans perceive our surroundings in an inherently multimodal way. However, the vast majority of data analysis idioms are exclusively visual, not using the apparent potential of combined designs. While the visualization field studies visual data analysis solutions, the sonification field investigates comparable solutions that convey data over non-speech audio. There are several similarities between the methods and design

* Editor / Organizer
† Editorial Assistant / Collector

theories of both approaches, such as the use of perceptual variables to encode data attributes, the use of marks and substrates to convey items, and the role of interaction in manipulating the data representations.

Over the last decades, both fields have established research communities, theoretical frameworks, and toolkit support. Although extensive research has been conducted on both auditory and visual representation of data, relatively little is known about their systematic and complementary combination for data exploration, analysis, and presentation. There are potential powerful synergies in combining both modalities to address the individual limitations of each other. Nevertheless, existing research on combinations has often focused only on one of the modalities.

This Dagstuhl Seminar aimed, first and foremost, at creating a space for discussion where the international communities of data visualization and data sonification came together to map and define a novel field for the improvement of human – data relationships. Over five days, 24 experts from both fields participated in group activities, inspirational sessions, keynote presentations, and informal gatherings. The seminar was the result of work conducted in the previous years by the organizers, including research projects, as well as a series of meetups and workshops and meetups of the AVAC – Audio Visual Analytics Community (`https://audio-visual-analytics.github.io`) at different venues:

- STAR and Panel @ EuroVis 2024 (Odense, Denmark): May 29-30, 2024
- Panel Discussion @ ICAD 2023 (Norrköping, Sweden): June 29, 2023
- Application Spotlight @ IEEE VIS (Oklahoma City, USA/hybrid): October 20, 2022
- 3rd Workshop on Audio-Visual Analytics @ Int. Conf. Advanced Visual Interfaces: June 7, 2022
- Workshop @ IEEE VIS (virtual): October 25, 2021
- Workshop @ Audio Mostly (virtual): September 3, 2021

Moreover, a State of the Art Report on data visualization and data sonification [1] was published as collaborative achievement of members of both communities. Based on these activities, the organizers identified a number of potential topics of discussion and shared them with the participants prior to the seminar. After an iterative, collective process of selection, we identified nine topics to be discussed in groups during the 5-day Dagstuhl Seminar:

- Design theory for joint visualization and sonification
- Design framework for joint visualization and sonification
- Application examples & success stories
- Tools and libraries
- Perception and cognition
- Redundancy and complementarity
- Interaction
- Accessibility
- Evaluation of joint visualization and sonification

During the seminar, after two inspirational keynote presentations that introduced the main achievements, techniques, tools, and challenges of both fields, participants were assigned to two topics of their choice, which they engaged in deeply in small discussion groups over the course of the week. Figure 1 shows the full schedule of the seminar.

The two *inspirational sessions* were designed as a light, engaging introduction for non-experts to visualization and sonification, and to provide a refreshing moment from the intense group discussions. In the evenings, participants gathered to showcase demos and current projects, exchange opinions on the respective fields, and overall enjoy each other's company. The excursion to Völklinger Hütte was an intriguing break that helped participants connect more deeply with the region that will provide long-lasting memories for the participants.

The work started during the seminar is far from being complete. However, we felt that great progress was made to bring the two perspectives closer together. The need to define a common idiom was a first challenge that became easier as the week progressed. Both communities made efforts to understand each other, which also involved self-reflection on one's field and open *eyes and ears* to the other's. Groups focused on mapping the borders of each field and began defining the key terms as a necessary preliminary step to build a new joint perspective. The sections of this document will report on each group's progress.

In a historical time when we as humans are dealing with the often overwhelming presence of data in our lives, and when researchers in the field carry the responsibility to facilitate sense-making of such data, this Dagstuhl Seminar has been widely recognised by participants as a milestone in establishing a common ground between the data visualization and data sonification community.

References

1 Kajetan Enge, Elias Elmquist, Valentina Caiola, Niklas Rönnberg, Alexander Rind, Michael Iber, Sara Lenzi, Fangfei Lan, Robert Höldrich, and Wolfgang Aigner. Open Your Ears and Take a Look: A State-of-the-Art Report on the Integration of Sonification and Visualization. *Computer Graphics Forum (EuroVis '24)*, 43(3):e15114, 2024.

Figure 1 Schedule of the seminar.

2 Table of Contents

3 Overview of Talks

On the first day of our seminar, we offered two talks and a self-introduction session, starting to build a bridge between the communities.

3.1 Sonification Background

Stephen Barrass (Sonification.com – Ainslie, AU)

This talk provided a background on sonification as a basis for sharing concepts and terminology with persons from the visualization community. At the beginning, some simple examples and a short historical timeline were presented. Next, canonical sonification techniques were introduced with examples. This was followed by an overview of more contemporary developments in techniques. Then, a final section introduced Sonic Information Design as an intentional approach using the *TaDa* Task-oriented Data Sensitive method for choosing a sonification technique.

3.2 Introduction to Information Visualization

Bongshin Lee (Yonsei University – Seoul, KR)

Information visualization is a powerful tool for understanding complex data by transforming it into visual representations that leverage human perception and cognition. In this talk, key concepts of information visualization were introduced for those new to the field, explaining why we visualize data and how it helps reveal patterns, trends, and insights. The role of interaction in enabling users to explore and manipulate data was discussed, making visualization a dynamic and engaging process. Additionally, the main purposes of information visualization, from exploratory data analysis to effective communication were covered. Finally, key challenges and opportunities in the field were highlighted.

3.3 Getting to know each other

Katharina Groß-Vogt (Universität für Musik & darstellende Kunst – Graz, AT)

First, fully embodied representation of the participants' backgrounds were enacted to sort along the alphabet (first names); along the time being in the research field; in 2D sorted on a world map; in 3D, stating your emotional feeling along valence, arousal, and dominance (lying on floor – standing). Then we conducted a speed dating on questions such as what is a good or bad example of sonification and visualization.

4 Inspiration Sessions

This Dagstuhl Seminar offered two *Inspiration Sessions*, one with a focus on audio and one focusing on visuals:

4.1 The Soundwalk

Kajetan Enge (Universität für Musik & darstellende Kunst – Graz, AT & St. Pölten University of Applied Sciences, AT)
Sara Lenzi (University of Deusto – Bilbao, ES & Ikerbasque Basque Foundation for Science – Bilbao, ES)

The "Soundwalk" is a practice born in the 1970s together with the emergence of soundscape studies [1, 2]. Soundscape is a term that refers to the totality of the sounds we hear in our acoustic environment, be it the sounds that we ourselves produce (breathing, walking) or sounds provoked by our interaction with the environment or by other agents in the close or distant world (e.g., cars, birds, airplanes...). Soundwalks are a tool that increases awareness of how we can explore and make sense of the world by actively listening to it. There are many ways to conduct a soundwalk. During this seminar, we selected and proposed to participants seven "Listening spots," both indoor and outdoor the Castle. Each Listening spot presented a suggestion for how to listen to the soundscape: "Imagine: Everything you hear is meant as a concert just for you," "Can you hear something that cannot be seen?", "How far can you hear?" ... Participants were encouraged to explore each listening spot on their own, in silence – some participants even found their own personal listening spots!

References

1 R. Murray Schafer. *The Soundscape. The Tuning of the World.* Knopfs, 1977.
2 Hildegard Westerkamp. Soundwalking. *Sound Heritage*, III(4), 1974.

4.2 Presentation and Activity on Alternative Designs

Jonathan C. Roberts (Bangor University, UK)

This session discussed design ideas to create many alternative designs. Starting with some background information on designs – looking briefly at an example by Jacques Bertin in his Semiology book – the session turned to getting participants in a group to consider alternative design ideas. The activity met the VisDice framework developed by Roberts and Owen [1], which used six inspirational images. In the activity, the image is chosen, and using lateral thinking, the people sketched a quick solution to visualize or sonify a dataset. The participants were split into four groups, each using the same test data. After sketching their own 2 cm x 2 cm quick sketch, they explained their idea to their peers. The next dice was rolled, giving a new dice image (inspirational image). The new idea must be different to the first. In this way the ideas change and are diverse from the initial idea. It helps to create new ideas. The process and ideology of generating alternatives were discussed, and the larger methodology of Five Design-Sheets [2] was also briefly presented.

References

1 Aron Owen and Jonathan C. Roberts. Inspire and create: Unveiling the potential of visdice in visualization design. In Nan Cao, Barbora Kozlikova, Jiazhi Xia, and Wesley Willett, editors, *IEEE VIS 2023 Posters: Visualization & Visual Analytics*. IEEE Computer Society Press, 2023.

2 Jonathan C. Roberts, Chris Headleand, and Panagiotis D. Ritsos. Sketching designs using the Five Design-Sheet methodology. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):419–428, 2015.

5 Working Groups

We structured our seminar in nine working groups.

5.1 Application Examples & Success Stories

Wolfgang Aigner (St. Pölten University of Applied Sciences, AT)
Stephen Barrass (Sonification.com – Ainslie, AU)
Johanna Schmidt (TU Wien, AT)
Kelly Snook (Kepler Concordia – Yamaguchi, JP)

The purpose of the *Application Examples & Success Stories* working group was to collect application examples and success stories for the combination of sonification and visualization.

5.1.1 Summary

The working group started the discussion with collecting examples of joint sonification and visualization, both in research and everyday situations. Especially when communicating the effectiveness of audio-visual feedback to people not familiar with sonification, the parking sensor in cars, metal detectors, and tools for localizing blockage in long pipes could be used as familiar examples. The working group continued the discussion by browsing the STAR report [1] and personal collections for audio-visual examples. From there, we identified key elements where sonification contributed to a data visualization. Afterward we looked into how visualization adds to sonification. A table was created that summarizes advantages for designers when using an integrated audio-visual approach. A list of application domains where audio-visual techniques are already used was created. The working group concluded the discussion by collecting examples of real-world situations where both audio and visual channels are required. Further developments could include the identification of real-world examples where both modalities are essential.

5.1.2 Application Areas

When looking at different audio-visual examples, we identified the following application areas where sonification and visualization are already used in combination:

- Astronomy
- Earth Science, Climate Change
- Physics
- Medicine and Health
- Molecular Science
- Biology, Zoology
- Social Sciences
- Economics
- Urban Studies
- Data Journalism

5.1.3 Results

We identified use cases and tasks where each modality (sonification and visualization) can profit from integration. Table 1 summarizes task and use cases where adding visual content to a sonification brings advantages to the analytics. Table 2 summarizes task and use cases where adding audio to a data visualization brings advantages to the analytics.

Table 1 The first column defines the task or use case that can be improved by adding visualization to a sonification. The second column describes the advantage for the potential users of the application.

Explore details	Audio can be used to set a context (e.g., bells give a comparison between different datasets/concepts). Visualization can then be used to explain the difference.
Exploration / Interaction	Visualization to steer the audio exploration of a dataset. Selecting details in visualization can help to focus on particular channels in the audio signal.
Orientation / Guidance	Point users to specific parts in the audio data which is interesting/important to listen to.

Table 2 The first column defines the task or use case that can be improved by adding audio to a visualization. The second column describes the advantage for the potential users of the application.

Sudden changes	We can hear sudden data changes better than we can see them. Rapid changes that are not shown very long (e.g., scrolling through a dataset) are hearable, but hardly visible and might be missed.
Small changes / Accentuation	Very small changes might not be visible (e.g., very little difference between bars), but these small changes would be hearable.
Repeating patterns	It is hard to see repeating patterns and compare them over time. This would work better when using audio.
Evaluation / Trust building	By not only watching a visualization, but also listening to a visualization, you can be sure that you did not miss anything and that you considered all small data changes. Audio may reveal uncertainty in a plot.
Different data scales	Listening to data from different data scales (e.g., tiny changes vs. large data scales) is possible, but visualizing data with different data scales is really hard.
More channels	With audio, it is possible to listen to more data channels than when watching the data (views get cluttered easily).
More foci	Possible to focus on a visualization (e.g., 3D representation) and hear a second parameter. Instead of switching between different views, audio can represent an additional channel (e.g., users select a data channel for audio and concentrate on another view).
More details	Different sound channels (e.g., for different data channels) can be heard and identified independently (if a dataset is constant or changing a lot). It is probably not possible to identify correlation.
Additional analysis channel	Adding audio can be compared to adding color to a visualization – it is then possible to represent, grasp and analyze more information at once.
Engagement	Sound/music can increase engagement, motivate people to watch a visualization, and help them understand the data. Users emerge while watching the visualization.
Increasing Trust	Sound helps to increase trust in automatic decisions of a machine (e.g., self-driving car, to indicate that it will turn or stop).
Feedback / Confirmation	When interacting with data, audio signals can be used for confirming the interaction (e.g., selection). Audio can be used to indicate the effect of an interaction (e.g., how many data points will be selected/removed by an interaction).
Transitions	Audio signals can prepare users for changes and transitions.

References

1 Kajetan Enge, Elias Elmquist, Valentina Caiola, Niklas Rönnberg, Alexander Rind, Michael Iber, Sara Lenzi, Fangfei Lan, Robert Höldrich, and Wolfgang Aigner. Open Your Ears and Take a Look: A State-of-the-Art Report on the Integration of Sonification and Visualization. *Computer Graphics Forum (EuroVis '24)*, 43(3):e15114, 2024.

5.2 Perception & Cognition

Elias Elmquist (Linköping University, SE)
Katharina Groß-Vogt (Universität für Musik & darstellende Kunst – Graz, AT)
Michael Krone (Hochschule für Technik – Stuttgart, DE)
Sita Vriend (Universität Stuttgart, DE)

We argue for using multi-modal displays as they meet best the nature of human perception, following an ecological approach of how our senses were shaped and how they interplay to heuristically build up a representation of our surroundings.

The properties of audio-visual perception vs. cognition, i.e., the lower vs. higher level processes of Gestalt psychology, have to be clear to allow for creating truly multi-modal displays. Therefore, we intend to recommend a set of testable guidelines, still on a hypothetic level, that motivate mapping choices in audio-visual displays:

If you follow these guidelines for creating an audio-visual representation, the cognitive load might still increase compared to the load of the purely visual or purely auditory representation, but the least increase possible. It might even decrease! And – in any case – the user will also gain more insights.

Guideline hypotheses

- **Utilize principles of grouping in gestalt psychology and of cross-modal perception for both auditory and visual elements in the audio-visual data representation.** For instance, pitch can enhance brightness perception. Many more examples need to be collected here.
- **The A+V representation needs to be redundant in at least one data dimension, for instance either via a spatial/temporal synchronization or linked via interaction.** This argument is supported by the real-world synchronization of audio-visual events. Furthermore we draw on Munzner's [1] argument on visualizations: it makes no sense to have exactly the same visualization of the same data; but also not to show a different visualization of two different data sets. The benefit is showing either two visualizations of the same data or representing different data by the same visualization.
  - Any dynamic sonification should involve a dynamic visualization (e.g., showing the position in the timeline).
  - An overview visualization should be linked to an overview sonification!
  - A visualization can be combined with interactive sonification (details on demand!)
- **If there is a fully technically redundant A+V representation, question if there is any added value.**
  - Technical redundancy: all data variables map to visual variables and to auditory variables.
  - "Communicative redundancy" [2]: one modality allows for different insights (e.g., emotions in sound, engagement). We need more research here.
- **Sound has to be synchronous or a little bit delayed (and never earlier than visuals).** We find evidence for this claim in the Temporal Ventriloquism effect [3], though we need to add more details here from psychophysics research.

- **We *should* always use spatial sounds.** The evidence for this is the spatial ventriloquism effect, and the fact that all visualization is spatial. We should make use of spatial sound rendering in order to better align A+V displays.
- **Be clear in creating either consonance OR dissonance on the level of indexicality between A and V representation.** Indexicality [4] is the level between iconic and symbolic representation.
- **A good A-V mapping involves to separately (!) make sure the polarity of the audio representation and visual representation correctly represent the change in data value (semantics).**
  - Mapping should consider audio-visual correspondence AND the a-data and the v-data mapping separately.
  - Scaling of A-V mappings and mappable differences AV-data
  - Dimension of A and V should be considered

We discussed a number of further possible guidelines, but these need more research, either through literature review or our own experiments:

- List experimental findings on experiences of how we interpret AV representations uniformly. Furthermore, we can learn AV representations – find more evidence from multi-modal learning experiments.
- Identify saliency principles to guide attention in AV representations, and find which auditory properties relate to visual ones in their being salient.

References

1 Tamara Munzner. *Visualization Analysis and Design.* AK Peters, Boca Raton, FL, USA, 2014.
2 Kajetan Enge, Elias Elmquist, Valentina Caiola, Niklas Rönnberg, Alexander Rind, Michael Iber, Sara Lenzi, Fangfei Lan, Robert Höldrich, and Wolfgang Aigner. Open Your Ears and Take a Look: A State-of-the-Art Report on the Integration of Sonification and Visualization. *Computer Graphics Forum (EuroVis '24)*, 43(3):e15114, 2024.
3 Daniel A. Slutsky and Gregg H. Recanzone. Temporal and spatial dependency of the ventriloquism effect. *NeuroReport*, 12(1):7, January 2001.
4 Dominic Oswald. Non-speech audio-semiotics: A review and revision of auditory icon and earcon theory. In *Proceedings of the 18th International Conference on Auditory Display (ICAD)*, Atlanta, GA, USA, June 2012. 18–21 June 2012.

5.3 Definition of a Research and Development Agenda for Integrated Audio-Visual Data Representation Tools and Libraries

Valentina Caiola (City University of Hong Kong, HK)
Thomas Hermann (Universität Bielefeld, DE)
Øystein Moseng (Highsoft AS – Vik I Sogn, NO)
Alexander Rind (St. Pölten University of Applied Sciences, AT)

We defined an audio-visual analytics idiom (AVAID), which is a systematic and reproducible audio-visual data representation. We followed a two-fold approach with a bottom-up phenomenological collection of current tools and libraries and a top-down scenario-based investigation with seven diverse personae. Outputs of the investigation:

a. Collection of categorization criteria for existing (visualization, sonification, and AV-integrated) tools, components and libraries
b. Identification of Interface Abstraction Levels – distributed on a continuum, where the two poles are (a) tools which are configured via a graphical user interface and (b) libraries/packages to be used for coding tools or components.
c. List of Requirements: List of needs and problems to be addressed by AVAIDs tools
d. Definition of minimal criteria for inclusion in AVAID tool set: **R**endering from **D**ata through a systematic representation **M**ethod (e.g., mapping, model) – DMR
e. Collection of Use Case Scenarios – along the interface abstraction levels continuum
f. Prototyping trade-offs

5.3.1 Defining the scope of tools and libraries

We defined **DMR** tools and libraries as complete systems that enable **R**endering from **D**ata through a systematic representation **M**ethod (e.g., mapping, model). We can differentiate between tools and libraries that facilitate this complete process, and components that support parts of this process. A full AVAID could for instance be setup as such:

- Data: a component to read/filter data (e.g., numpy for loading csv files),
- Mapping/Model: a component for mapping data (e.g., a python function to create a Csound score file),
- Rendering (Auditory Representation): CSound is a sound rendering component usable for sonifications that focuses exclusively on sound rendering (R).
- Rendering (Visual Representation): for instance the data could be plotted using scalable vector graphics (SVG) in the web browser.

5.3.2 Mapping the landscape of AVAID tools and libraries

We identified the following dimensions as useful for differentiating between AVAID tools and libraries:

- Platform (e.g., web, native, specialized)
- Environment (e.g., jupyter notebooks, IDEs)
- Self-contained DMR tools vs component tools vs. libraries
- Interface abstraction level (e.g., GUI, text, code)
- Popularity
- Licensing (open/closed)
- Learning curve vs. perceived value (“Return on time invested”)
- Capabilities / versatility
- Data type support (e.g., map, graph, tabular)
- Data input/output, import/export
- Output capabilities
- Static vs dynamic output (interactivity)
- Visual rendering types (pixel graphics, vector graphics)
- Rendered audio types (sampled audio vs. score / higher level representation)
- Synchronization (e.g., between auditory and visual rendering)
- Collaborative editing or production features
- Extensibility

We believe the AVAID tooling space will need to cover a spectrum of different interface abstraction levels and levels of capabilities. This is analogous to the existing tooling space in visualization. Observing the current tools and libraries state of the art, we identified the need for audio-visual tools to develop AVAIDs, and came up with a narrow research question: RQ1: "What are the needs/problems that should be solved by AVAID tools across different interface abstraction levels?"

- functionality (services) matter more than rendering systems
- suitable rendering abstractions (to avoid engine specific lock-in)
- shareability/extensibility to promote community (presets, templates, extensions)
- enable (best) integration of the visualization and sonification modalities (synchronization, Multiple Coordinated Views (MCV) between modalities,...)
- Input/Output – Import/Export – Formats
- Users lacking knowledge of visualization and sonification – good output anyway – capture intent and context
- real-time rendering for enhanced interactivity
- continuum of flexibility versatility in data handling, method, and rendering
- data types (map, graph, tabular) and data size (performance)
- production speed (prototyping)
- collaborative production

5.3.3 Prototyping

We discussed how prototyping relates to AVAIDs and libraries and identified three dimensions that form a triangle of trade-offs:

- precision (or fidelity) (e.g., mock or subsampled data, imprecise representation method, or sketchy rendering)
- aesthetics
- speed (time needed to create prototype & products)

We acknowledge the speed quality trade-off in prototyping, speed referring to inverse time needed to create the prototype/system. Along the quality direction we usually face another trade-off, the one between precision and aesthetics, where the former refers to how accurately information is conveyed to the users and the latter refers to how pleasant and aesthetically the media is judged. Altogether, a ternary "prototyping space" emerges in which different AVAIDs can be positioned. AVAIDs may serve as an enabler of this process through their capabilities.

5.4 Towards a Design Framework

Kajetan Enge (Universität für Musik & darstellende Kunst – Graz, AT & St. Pölten University of Applied Sciences, AT)
Dietmar Offenhuber (Northeastern University – Boston, US)
Christina Gillmann (Fraunhofer FIT – St. Augustin, DE)
Bruce Walker (Georgia Institute of Technology – Atlanta, US)
Sara Lenzi (University of Deusto – Bilbao, ES & Ikerbasque Basque Foundation for Science – Bilbao, ES)

Creating audio-visual artifacts requires design. And design is hard, especially in an interdisciplinary setting. Nevertheless, Design as a discipline can provide the principles, processes, and tools that our field needs to produce impactful outcomes. Our main message is to be intentional, systematic, and interdisciplinary. This field will need to define its design patterns, best practices, and implementation strategies that achieve the design goals.

The design process as we see it starts with clearly specified Design Goals. These include hypotheses, purposes, intended insights, communicative intention, and potential outcomes. Understanding the ecosystem in which the AVA exists is crucial. The ecosystem is shaped by constraints, previous designs, and also future possibilities. The AVA Design Space is vast, requiring many decisions along multiple dimensions. Design Patterns can provide a process to navigate through the design space. Design Patterns can reflect the values, approaches, and best practices of our community. Finally, implementation can leverage all that is known about perception, cognition, aesthetics, communication, emotion, technology, etc., reflecting the depth of interdisciplinarity in the field.

5.5 Interaction

Mohammad Ghoniem (Luxembourg Institute of Science and Technology, LU)
Bongshin Lee (Yonsei University – Seoul, KR)
Sandra Pauletto (KTH Royal Institute of Technology – Stockholm, SE)
Jonathan C. Roberts (Bangor University, UK)
Niklas Rönnberg (Linköping University, SE)
Jonathan Zong (MIT – Cambridge, US)

When designing an interactive audio-visual system, we begin with data as the foundation. Our goal is to present information in a way that facilitates people to understand, learn, or engage in an enjoyable experience. The data itself shapes the auditory and visual experience, serving as the foundation of interaction.

Interaction enables individuals or groups to engage with the system, adapt their experience, and control the data exploration, and raises key questions: How is this system setup? What principles are there when setting up the system? How can we control the system and what actions do we enact upon the system? What devices are there to control the system? How do we display the output, and where is it displayed?

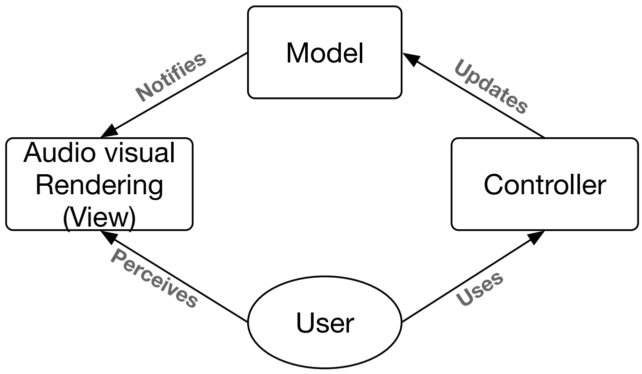

Figure 2 Model view controller for audio-visual renderings.

The model-view-controller (MVC) pattern, widely used in human-computer interaction [1, 2], offers a valuable framework for structuring these experiences (see Figure 2). In this context, we suggest using the term "rendering" instead of "view" because "view" carries a strong visual association. In contrast, both sound and visuals can be rendered, making "rendering" a more inclusive term for this discussion.

The *user* wants to understand the information, hear the data, and experience the visual feedback. There is an intent, an aim for their task. The data underpins the system. It can come from a sensor of the world (the distance from the car bumper to the tree) or from a questionnaire or survey (how happy are people in the world). Before it is rendered in visual or auditory form, it may need to go through cleaning, preparation and transformation steps to be ready for the audio/visual mapping. The result of this process is a *data model*. It is this data model that will be turned into sound and visualised. The user then interacts via a *controller* to express their intents, or goals, by controlling aspects or parts of the model that are rendered. The controller can be a device, an object of various shapes and materials, or more traditionally a slider on a user interface, button on a keyboard, or even their body in the world. The data-model now needs to be rendered – viewed and heard in some form. The *"rendering"* is a visual or auditory presentation of the data model, based on an output device, for example, speakers or screens in the room, or speakers or screens embedded in the objects that are interacted with, on headphones or in an immersed head-mounted display, mixed along with the visuals on a screen, handheld devices such as smartphone, tablet or small display such as a watch.

These ideas can be expanded and discussed further. Each of the individual parts, the model, rendering, controller and role of the user can be explored further. One important focus is that many instances of each part of the architecture are relevant. They are necessarily coordinated to provide a full and rich user experience. Multiple users can be in the system, interacting together, and controlling the audio-visual experience. Many renderings support multiple views or alternative presentations of the data. Various types of controllers enable different commands – some offering more intuitive interfaces, others designed for precise control, and some focused on enhancing user enjoyment. In addition, different controls, mappings, and experiences cater to specific needs, ensuring accessibility. For example, auditory feedback allows visually impaired users to engage with the system. People can also learn better when they can explore multiple and alternative ways of experiencing or interacting with the data.

References
1 John Hunt. The Model-View-Controller Architecture. In *Smalltalk and Object Orientation*, pages 266–278. Springer, London, 1997.
2 John Zukowski. The Model-View-Controller Architecture. In *John Zukowski's Definitive Guide to Swing for Java 2*, pages 71–77. Apress, Berkeley, CA, 1999.

5.6 Accessibility

Bongshin Lee (Yonsei University – Seoul, KR)
Øystein Moseng (Highsoft AS – Vik I Sogn, NO)
Jonathan C. Roberts (Bangor University, UK)
Bruce Walker (Georgia Institute of Technology – Atlanta, US)
Jonathan Zong (MIT – Cambridge, US)

The accessibility working group explored the role of multimodal audio-visual interfaces in improving accessibility, focusing on sonification and visualization. Discussions centered on the relationship between accessibility, inclusion, and usability, emphasizing the need for supporting diverse literacy levels and cognitive abilities. A key theme was redundancy across multiple modalities – how different representations can enable insights even if full redundancy across modalities is unattainable. We examined the balance between alternative and redundant representations, considering how auditory and visual modalities can enhance understanding without assuming that increased information bandwidth improves accessibility. The group also discussed strategies for achieving meaningful redundancy, including comparisons between perceptual mappings. We considered how interactive techniques can bridge gaps and link modalities – such as visually inspecting data while controlling sonification in time. The working group continues to discuss future research directions, highlighting the need for practical examples, interactive approaches, and frameworks that support accessible multimodal data representation.

5.7 Redundancy and Complementarity

Wolfgang Aigner (St. Pölten University of Applied Sciences, AT)
Sandra Pauletto (KTH Royal Institute of Technology – Stockholm, SE)
Elias Elmquist (Linköping University, SE)
Kelly Snook (Kepler Concordia – Yamaguchi, JP)
Katharina Groß-Vogt (Universität für Musik & darstellende Kunst – Graz, AT)
Christina Gillmann (Fraunhofer FIT – St. Augustin, DE)

We developed the beginning of a redundancy-complementarity model for AV representations. Through discussions, which started with individual examples and understandings of redundancy/complementarity, we realised that redundancy/complementarity can be defined in the data, representation, and task domains. A **fully redundant** representation involves both representations (A and V) visualising/sonifying the same dataset and the user being able to do the same task with the audio-only, or visual-only or audio-visual representations. A **semi-redundant** representation might utilize two (or more) overlapping datasets, some visualised and others sonified. In this case, the sonification affords a task and the visualisation might afford a different task. A **fully complementary** representation involves separately sonifying and visualising two (or more) different datasets. These representations afford different tasks, but the fact that they are happening at the same time might make emerge an unexpected meaning that could be then used in some way, for example to create an art work. Furthermore, the group summarised possible benefits of these representations. Note that the following list of benefits is not exhaustive. A fully redundant representation might afford inclusiveness, the possibility of switching modality, safety and trust. The semi-redundant representation can afford engagement, reduction of cognitive load, representation of multiple dimensions (e.g., uncluttering), multi-tasking, richer understandings and trust. The fully complementary representation might afford the serendipitous emergence of a new/unexpected meaning, which could be beneficial in some contexts (for example an artistic context).

5.8 Towards a Design Theory

Kajetan Enge (Universität für Musik & darstellende Kunst – Graz, AT & St. Pölten University of Applied Sciences, AT)
Mohammad Ghoniem (Luxembourg Institute of Science and Technology, LU)
Thomas Hermann (Universität Bielefeld, DE)
Sara Lenzi (University of Deusto – Bilbao, ES & Ikerbasque Basque Foundation for Science – Bilbao, ES)
Niklas Rönnberg (Linköping University, SE)
Johanna Schmidt (TU Wien, AT)

The working group based its discussion on the Design Framework suggested by the earlier discussion group. We think of principles of Audio-Visual Idioms as encapsulating the Design Framework into a Design Theory. Four concerns that were identified are:

1. **Intention** – Intentionality: We suggest that in the context of audio-visual design, Intention refers to the deliberate integration of sonification and visualization by the designer in order to support sensemaking, presentation, and storytelling about the data.
2. **Integration** – different levels of integration (low → high):
   a. Hardware integration of output/presentation devices (e.g., speakers, headphones, screens).
   b. Technical integration: implementation of (combining) two modalities.
   c. Audio-visual integration consists in the coordination of visual and auditory renderings as they depend on the same underlying data (MVC Model View Controller Architecture [1, 2]).
   d. Sensory integration: integration of two (or more) perceptual stimuli. This is what comes from the real world, e.g., we hear and see a dog at the same time. This has to do with what our perception is exposed to. Sensorial perception is a requisite.

e. Cognitive/mental integration: it happens when a person understands the stimuli as coming from the same source. This association might be only imagined, a concept in our mind (vs sensory integration).

3. **Sensemaking**: We expect our audience to take an analytical perspective of our artifact (i.e., audio-visual artifact) in order to make sense of the phenomenon of interest.
4. **Artifact**: In our context, an artifact implies an audio-visual representation of a dataset. An audio-visual idiom (grammar) without data values is not an artifact.

Work on Design Theory will consist in identifying principles (such as unified constructs or a unified multiple coordinated AV-renderings framework) ensuring a deeper integration of the two modalities.

References

1 John Hunt. The Model-View-Controller Architecture. In *Smalltalk and Object Orientation*, pages 266–278. Springer, London, 1997.
2 John Zukowski. The Model-View-Controller Architecture. In *John Zukowski's Definitive Guide to Swing for Java 2*, pages 71–77. Apress, Berkeley, CA, 1999.

5.9 Evaluation

Stephen Barrass (sonification.com – Ainslie, AU)
Valentina Caiola (City University of Hong Kong, HK)
Michael Krone (Hochschule für Technik – Stuttgart, DE)
Alexander Rind (St. Pölten University of Applied Sciences, AT)
Sita Vriend (Universität Stuttgart, DE)

The purpose of this working group was to discuss evaluation of audio-visual (AV) idioms. First, we discussed the necessity for evaluation of AV idioms. Then, we identified evaluation methods from visualization, sonification, and human-computer interaction (HCI) and how they can be arranged into a decision map as guidance for researchers and practitioners. We discussed the special considerations for evaluating audio-visual idioms: auditory impairments, sound literacy (e.g., musician/non-musician), technical test conditions (e.g., headphones/ speakers, room setup). Furthermore, the way how audio-visual idioms are not fully redundant makes it necessary to evaluate such idioms in an integrated way rather than visual separately from sound.

5.9.1 Why Evaluate AV Idioms?

We first discussed whether evaluation of AV idioms is always necessary. Some projects that involve AV idioms may not need to be evaluated. For example, the exploration of the design space and novel techniques may inherently be valuable and not need to be evaluated. The fields of sonification and visualization exhibit different trends in evaluating their research. While only few sonification papers include an evaluation [4], different forms of evaluations are widespread in visualization, which is (partly) due to publication venues often requiring an evaluation in order to publish. Of research on audio-visual data representations about two third include an evaluation [6]. Evaluation has the potential to reveal how an AV idiom is useful. Hence, we urge researchers to consider evaluation as part of their process, not merely as a formality.

5.9.2 Evaluation Frameworks and Methods

There is currently no framework or method for the evaluation of AV idioms specifically. Instead, we identified evaluation methods from visualization, sonification, and HCI. These methods can be used and adapted to suit the evaluation of AV idioms.

First, we like to point out evaluation methods from HCI. Evaluation methods from HCI are typically user-centered, meaning that the user is central to the design process. HCI distinguishes between formative and summative evaluation methods. Formative evaluation methods on the one hand are considered from the beginning and takes an iterative approach to the design of an artifact [27]. Visualization research has adopted formative HCI methods in visualization design studies [19, 25]. Summative evaluation methods on the other hand, are conducted at the end of a development cycle and tests how well a design performs. The design is often compared against benchmarks. Research in sonification and visualization often evaluates performance measures such as reaction time and error rate [15], but learnability [2], working memory load [22] and psychophysics [8, 23] has also been investigated.

From HCI and related fields, we identified the following evaluation techniques that would be useful for researcher to explore for the evaluation of AV idioms: (1) Empirical evaluation, (2) heuristic evaluation, (3) third-wave HCI methodologies, and (4) alternative methodologies.

1. **Empirical evaluation techniques** rely on experimental procedures to answer research questions. Experimental procedures can collect and analyze qualitative data such as in grounded theory [16], or quantitative data as long as the experiment is replicable. Validation of a data representation by comparing it to a second representation typically are empirical as well. Empirical evaluation techniques are widely used and accepted in visualization [17] and have been applied to sonification as well. Measures used in empirical evaluation vary widely and depend on the aspect a researcher would like to investigate. Some examples include standardized questionnaires such as the NASA-TLX on perceived workload [14], visualization specific questionnaires [10], questionnaires assessing usability of AV idioms [24], the visualization literacy questionnaire [18], among others. Visualization research has a rich tradition of mainly empirical evaluation techniques that could be adapted to evaluate AV idioms. For an introduction we could like to point the reader to *An Introduction and Guide to Evaluation of Visualization Techniques Through User Studies* [11] and *Patterns for visualization evaluation* [5].
2. **Heuristic evaluation** can identify usability issues in an artifact throughout the design cycles [20]. A heuristics evaluation involves a group of experts that examine and judge an artifact, often according to a list of principles. In HCI, many heuristic principles exist [21]. However, specific visualization heuristics have been developed [9]. We think that findings from psychoaccoustics and visual perception research could inform the design of AV idioms and could act as heuristics. There is an opportunity for future research to identify and formulate heuristics for AV idioms specifically.
3. **Third-wave HCI methodologies** go beyond the user-technology relationship and recognizes that interaction is a form of meaning creation. Methodologies thus focus on experiences and meaning-making in a broader context, beyond the artifact and its user [3, 13]. One such methodology is phenomenology: "a philosophical movement that seeks to analyze the relations between human beings and their world rather than as a method for describing reality" [26].
4. Lastly, we would like to point out **alternative evaluation methods**. We think such methods could be valuable to evaluate AV idioms too. For example: critical self-reflection, which is used in humanities [1, 7].

5.9.3 Special Considerations of the Evaluation of AV Idioms

Considering evaluation methods from adjacent fields is a good start. Such methods can be adapted to evaluate AV idioms. However, evaluations should be tailored to specific research goals. Hence, we identified special consideration for the evaluations of AV idioms. We came up with the following decision checklist to guide the evaluation of AV idioms:

1. **State the intent** of the AV idiom and possible hypotheses and assumptions: A design is usually created with intent. This intent should be testable. Additionally, the researcher should state their assumptions and expectations. These can be formulated as testable hypotheses.
2. **Define specificities:** Define who the intended user of the AV idioms is, what task(s) they will perform with it and in what context they will use it.
3. **Summative vs. Formative:** A researcher should decide how they intend to evaluate their AV idiom before they start design. Do they intend to involve the user throughout development? Then formative evaluation methods will apply. Or will the researcher compare the AV idiom once it is created? Then summative methods are better suited.
4. **Define "good":** Researchers might want to evaluate if their AV idiom is better than an alternative representation. In such a case it is important to define what it means for the AV idiom to be good. What good means depends on the intent and specificities of the designed AV idiom. For example, should the designed AV idiom lead to fewer errors or should the designed AV idiom be more memorable.
5. **Choose methodology and method:** The researcher should choose a suitable method based on the previous points. In both visualization and sonification, empirical methodology is often chosen. Empirical methods are especially suitable for summative approaches. This methodology include both quantitative and qualitative methods that can be used to validate a design. However, we also want to point to alternative methodologies from other fields. For example critical self-reflection which is commonly used in humanities, phenomenology which is widely accepted in design, and third-wave HCI methodologies.
6. **Choose measures:** What to measure should be chosen based previous points, especially on the methods chosen and what "good" means to the researcher. For example, a researcher wants to know if their AV idiom conveys the same information to sighted, deaf and blind users. In such a case, the researcher might chose a qualitative method to measure the users' comprehension of the AV idiom.

We recommend that researchers plan how to evaluate their AV idiom before they begin development. An AV idiom is often designed with a specific intent. This intent needs to be stated, as well assumptions and hypotheses. How to evaluate is guided by these intentions, assumptions and/or hypotheses. For example, an AV idiom is designed with the intent to help children learn about space. The evaluation should then test the learnability of the AV idiom with children.

Furthermore, design often requires specificity. AV idiom are often developed with a specific user in mind who will use the design in a specific context to perform determined tasks. These specificities need to not only be considered when designing the AV idiom but also when planning the evaluation. In the case of a summative evaluation, especially when looking to validate an AV idiom, a baseline must be chosen when evaluating an AV idiom. We concluded that AV idioms should be tested as an integrated system when exploring the design space of AV idioms. Hence, as a general rule of thumb, AV idioms should only be compared to alternative AV idioms. This is because Gibsodian perception [12], considers human perception as an active integrated system, rather than as passive and separate channels of sensation. This theory of ecological perception has enormous implications for the evaluation of multisensory interfaces! The whole is more than the sum of the parts. Evaluating channels separately will not provide a true description of the whole system. Furthermore the senses provide different kinds of information and a simply redundant interface is the least likely to provide perceptual benefits and advantages. Hence, it is not useful to compare visualization to AV idioms. However, we identified exceptions. For example, a researcher might compare an AV idiom to a visualization if they want to explore if the combination with sonification adds something useful to the visualization. Or if an AV idiom intends to be accessible for blind, sighted, and deaf individuals.

There are, of course, further factors that will influence the results of an evaluation. On the one hand, the background and personal factors of the user will play a role (e.g., musical training or auditory impairments). On the other hand, the technical setup will play a role (e.g., room/lab settings and the output device). Therefore, such factors need to be documented correctly and thoroughly considered when drawing conclusions based on the evaluation results.

However, we see the need for evaluation methods specifically developed for AV idioms. Further research should identify, adapt and extend evaluation methods from HCI, visualization and sonification useful for the evaluation of AV idioms. For example, how do we adapt evaluation methods to evaluate fully redundant AV idioms, fully complementary AV idioms, or mixed redundant AV idioms?

References

1 J Dewey. *How we think: a restatement of the relation of reflective thinking to the educative process.* D.C. Heath & Co Publishers, 1933.

2 Tilman Dingler, Jeffrey Lindsay, and Bruce N Walker. Learnabiltiy of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In *Proceedings of the 14th international conference on auditory display.* International Community for Auditory Display, 2008.

3 Emanuel Felipe Duarte and M. Cecilia C. Baranauskas. Revisiting the three HCI waves: a preliminary discussion on philosophy of science and research paradigms. In *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems*, pages 38:1–38:4, New York, NY, USA, 2016. ACM.

4 Gaël Dubus and Roberto Bresin. A systematic review of mapping strategies for the sonification of physical quantities. *PloS one*, 8(12):e82491, 2013.

5 Niklas Elmqvist and Ji Soo Yi. Patterns for visualization evaluation. In *Proceedings of the 2012 BELIV workshop: Beyond time and errors – novel evaluation methods for visualization*, Beliv '12, pages 12:1–12:8, New York, NY, USA, 2012. Association for Computing Machinery.

6 Kajetan Enge, Elias Elmquist, Valentina Caiola, Niklas Rönnberg, Alexander Rind, Michael Iber, Sara Lenzi, Fangfei Lan, Robert Höldrich, and Wolfgang Aigner. Open Your Ears and Take a Look: A State-of-the-Art Report on the Integration of Sonification and Visualization. *Computer Graphics Forum (EuroVis '24)*, 43(3):e15114, 2024.

7 Janet Eyler, Dwight E Giles, and Angela Schmeide. *A practitioner's guide to reflection in service-learning: Student voices & reflections.* Vanderbilt University, 1996.

8 Jamie Ferguson and Stephen A. Brewster. Evaluation of psychoacoustic sound parameters for sonification. In *Proceedings of the 19th ACM international conference on multimodal interaction*, Icmi '17, pages 120–127, New York, NY, USA, 2017. Association for Computing Machinery.

9 Camilla Forsell. Evaluation in Information Visualization: Heuristic Evaluation. In *2012 16th International Conference on Information Visualisation*, pages 136–142, 2012. tex.ids= forsell_2012_evaluation ISSN: 2375-0138.

10 Camilla Forsell and Matthew Cooper. Questionnaires for evaluation in information visualization. In *Proceedings of the 2012 BELIV workshop: Beyond time and errors – novel evaluation methods for visualization*, Beliv '12, pages 16:1–16:3, New York, NY, USA, 2012. ACM.
11 Camilla Forsell and Matthew Cooper. An Introduction and Guide to Evaluation of Visualization Techniques Through User Studies. In Weidong Huang, editor, *Handbook of Human Centric Visualization*, pages 285–313. Springer, New York, NY, 2014.
12 James Jerome Gibson. *The senses considered as perceptual systems.* Bloomsbury Academic, 1966.
13 Steve Harrison, Deborah Tatar, and Phoebe Sengers. The three paradigms of HCI. In *Alt. Chi. Session at the SIGCHI Conference on human factors in computing systems San Jose, California, USA*, pages 1–18, 2007.
14 Sandra G. Hart. NASA-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, October 2006.
15 Jessica Hullman, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. In pursuit of error: a survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):903–913, 2019.
16 Petra Isenberg, Torre Zuk, Christopher Collins, and Sheelagh Carpendale. Grounded evaluation of information visualizations. In *Proceedings of the 2008 workshop on beyond time and errors: Novel EvaLuation methods for information visualization*, Beliv '08, pages 6:1–6:8, New York, NY, USA, 2008. Association for Computing Machinery.
17 Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012.
18 Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. VLAT: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):551–560, 2017.
19 Miriah Meyer and Jason Dykes. Criteria for Rigor in Visualization Design Study. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):87–97, January 2020. tex.ids= Meyer2020designstudy conferenceName: IEEE Transactions on Visualization and Computer Graphics.
20 Jakob Nielsen. How to conduct a heuristic evaluation. *Nielsen Norman Group*, 1(1):8, 1995.
21 Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256, New York, NY, USA, 1990. ACM.
22 Lace M.K. Padilla, Spencer C. Castro, P. Samuel Quinan, Ian T. Ruginski, and Sarah H. Creem-Regehr. Toward objective evaluation of working memory in visualizations: a case study using pupillometry and a dual-task paradigm. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):332–342, 2020.
23 Ghulam Jilani Quadri. Toward constructing frameworks for task- and design-optimized visualizations. *IEEE Computer Graphics and Applications*, 44(5):104–113, 2024.
24 Niklas Rönnberg and Camilla Forsell. Questionnaires assessing usability of audio-visual representations. In *AVI 2022 Workshop on Audio-Visual Analytics (WAVA22)*, Frascati, Rome, Italy, May 2022. Zenodo.
25 Michael Sedlmair, Miriah Meyer, and Tamara Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.
26 Peter-Paul Verbeek. *Moralizing technology: Understanding and designing the morality of things.* University of Chicago press, 2011.

27 John Zimmerman, Jodi Forlizzi, and Shelley Evenson. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 493–502, New York, NY, USA, April 2007. Association for Computing Machinery.

6 Open Mic Sessions

In our seminar, we held Open Mic Sessions on Monday, Tuesday, and Thursday evenings in the castle's wine cellar. Mainly, the following topics and projects were presented and discussed:

- Jonathan Zong presented his work on *Umwelt* [2].
- Kajetan Enge presented *theoretical constructs* bridging the theory-gap between the sonification and visualization fields [1].
- Kelly Snook presented her MiMu Gloves[1], a pair of gloves for versatile music composing and performing trough gestural movements.

References

1 Kajetan Enge, Alexander Rind, Michael Iber, Robert Höldrich, and Wolfgang Aigner. Towards a unified terminology for sonification and visualization. *Personal and Ubiquitous Computing*, 27(5):1949–1963, oct 2023.

2 Jonathan Zong, Isabella Pedraza Pineros, Mengzhu (Katie) Chen, Daniel Hajas, and Arvind Satyanarayan. Umwelt: Accessible Structured Editing of Multi-Modal Data Representations. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, pages 1–20, New York, NY, USA, may 2024. Association for Computing Machinery.

[1] `https://mimugloves.com/`

Participants

- Wolfgang Aigner
FH – St. Pölten, AT
- Stephen Barrass
Sonification – Ainslie, AU
- Valentina Caiola
City University –
Hong Kong, HK
- Elias Elmquist
Linköping University, SE
- Kajetan Enge
Universität für Musik &
darstellende Kunst – Graz, AT
- Mohammad Ghoniem
Luxembourg Institute of Science
& Technology, LU
- Christina Gillmann
Fraunhofer FIT –
Sankt Augustin, DE
- Katharina Groß-Vogt
Universität für Musik &
darstellende Kunst – Graz, AT
- Thomas Hermann
Universität Bielefeld, DE
- Michael Krone
Hochschule für Technik –
Stuttgart, DE
- Bongshin Lee
Yonsei University – Seoul, KR
- Sara Lenzi
University of Deusto – Bilbao, ES
& Ikerbasque Basque Foundation
for Science – Bilbao, ES
- Øystein Moseng
Highsoft AS – Vik I Sogn, NO
- Dietmar Offenhuber
Northeastern University –
Boston, US
- Sandra Pauletto
KTH Royal Institute of
Technology – Stockholm, SE
- Alexander Rind
FH – St. Pölten, AT
- Jonathan C. Roberts
Bangor University, GB
- Niklas Rönnberg
Linköping University, SE
- Johanna Schmidt
TU Wien, AT
- Kelly Snook
Kepler Concordia –
Yamaguchi, JP
- Sita Vriend
Universität Stuttgart, DE
- Bruce N. Walker
Georgia Institute of Technology –
Atlanta, US
- Jonathan Zong
MIT – Cambridge, US

Report from Dagstuhl Seminar 25081

Semirings in Databases, Automata, and Logic

Guillermo Badia*[1], Manfred Droste*[2], Phokion G. Kolaitis*[3], Carles Noguera*[4], Sophie Brinke†[5], Lovro Mrkonjić†[6], and Gaia Petreni†[7]

1 **University of Queensland – Brisbane, AU.** g.badia@uq.edu.au
2 **Universität Leipzig, DE.** droste@informatik.uni-leipzig.de
3 **University of California – Santa Cruz, US & IBM Almaden Research Center – San Jose, US.** kolaitis@ucsc.edu
4 **University of Siena, IT.** carles.noguera@unisi.it
5 **RWTH Aachen, DE.** brinke@logic.rwth-aachen.de
6 **RWTH Aachen, DE.** mrkonjiclovro@gmail.com
7 **University of Siena, IT.** gai.petreni@student.unisi.it

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 25081 "Semirings in Databases, Automata, and Logic", which was held from February 16 to 21, 2025. The seminar focused on semirings, a class of algebraic structures with many applications in computer science, particularly in databases and automata. Semirings are used in databases to annotate tuples in the input and output relations of queries (in particular, in the case of bag semantics, using the semiring of natural numbers) allowing to model several relevant aspects of databases. In automata theory, semirings allow to define weighted automata, which have applications in natural language processing, speech recognition, and image compression. Moreover, semirings are strongly related to the algebraic semantics of many-valued logics. The seminar brought together researchers from the communities mentioned above, and it developed a research agenda for studying semirings, guided by a collection of diverse applications. This led to several new collaborations between members from different communities, including joint work for publications.

Seminar February 16–21, 2025 – https://www.dagstuhl.de/25081

2012 ACM Subject Classification Theory of computation → Database theory; Theory of computation → Formal languages and automata theory; Theory of computation → Logic; Mathematics of computing → Discrete mathematics

Keywords and phrases databases, finite model theory, multi-valued logic, provenance, semirings, weighted automata, weighted logic

Digital Object Identifier 10.4230/DagRep.15.2.89

* Editor / Organizer
† Editorial Assistant / Collector

Semirings in Databases, Automata, and Logic, *Dagstuhl Reports*, Vol. 15, Issue 2, pp. 89–109
Editors: Guillermo Badia, Manfred Droste, Phokion G. Kolaitis, Carles Noguera, Sophie Brinke, Lovro Mrkonjić, and Gaia Petreni
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Guillermo Badia (University of Queensland – Brisbane, AU)
Manfred Droste (Universität Leipzig, DE)
Phokion G. Kolaitis (University of California – Santa Cruz, US & IBM Almaden Research Center – San Jose, US)
Carles Noguera (University of Siena, IT)

Semirings are fundamental algebraic structures that in recent times have found a number of applications to computer science, especially in the areas of databases and automata. On the side of databases, commercial query languages, such as SQL, use bag semantics, instead of set semantics, to evaluate relational database queries, which means that the semiring of the natural numbers is used to annotate tuples in the input and output relations. More generally, the annotations can be values in some fixed semiring; this gives a common generalization of both set semantics and bag semantics of database queries, and also makes it possible to model other situations in which one is interested, e.g., in the probability or the reliability of an answer. Furthermore, semirings of polynomials have been successfully used to carry out a rigorous study of provenance in databases. On the side of automata, semirings are used to define weighted automata, which are nondeterministic finite automata augmented with values from a semiring as weights on the transitions. These weights may model, e.g., the cost involved when executing a transition, the amount of resources or time needed for this, or the probability or reliability of its successful execution. Weighted automata have found numerous applications to natural language processing, speech recognition, and algorithms for digital image compression.

These applications have inspired numerous investigations in the logic-in-computer-science community. For instance, motivated by the work on semirings in databases, the evaluation of arbitrary first-order formulas in semirings has been recently explored in the literature, where, among other things, the relation between elementary equivalence (i.e., indistinguishability in first-order logic) and isomorphism on finite structures has been examined. In the same vein, various notions of locality in the semiring framework have been studied and zero-one laws and convergence laws have been obtained, whereas Ehrenfeucht–Fraïssé games have been used to characterize equivalence up to bounded quantifier rank under semiring semantics. As regards semirings and automata, Kleene's classical theorem on the characterization of languages recognized by automata with languages denoted by regular expressions was extended by Schützenberger to the behaviors of weighted automata and rational power series. By a fundamental theorem of Büchi, Elgot, and Trakhtenbrot, finite automata have the same expressive power on words as monadic second-order logic; therefore, various problems about this logic on words are decidable. In recent years, a suitable weighted logic was developed and shown to have the same expressive power on words, trees, and graphs as weighted automata. Consequently, this weighted logic has similar decidability properties on these structures as the unweighted monadic second-order logic. In turn, both of these approaches could be seen as part of the framework of many-valued logics, where algebraic structures expanding semirings with suitable operations have been studied for a long time.

This seminar brought together researchers from three communities that share an interest in semirings, namely databases, automata and multi-valued logic. The goal was to develop a research agenda, featuring tutorials, contributed talks, and discussions on open problems and future directions. Overall, the seminar was a success.

Organization of the Seminar. The seminar was held between February 16–21, 2025 from Monday to Friday. There were five invited tutorials with long 75-minute talks, two from the database community, two from the automata community, and one from the multi-valued logic community, that helped define a common language and a common set of problems, and 16 contributed 25-minute talks from experts in all these fields. The tutorials were given by

- Val Tannen on the semiring framework of data(base) provenance;
- Erik Paul on weighted automata and weighted logics;
- Erich Grädel on the model theory of semiring semantics;
- Sara Ugolini on the algebraic semantics of many-valued logics; and
- Peter Kostolányi on models of computation over semirings.

In addition, there was a session where five open problems (some of them rather major) were presented. Our collectors, Sophie Brinke, Lovro Mrkonjić, and Gaia Petreni, recorded all open problems, and later collected them for inclusion in this report. Finally, there was a closing session to formulate future research directions. The schedule was developed in a way that allowed sufficient time for informal discussions and exchange of ideas between participants.

Outcomes of the Seminar. There are several major outcomes:

- The presence of participants with very diverse backgrounds enabled us to exchange interesting ideas. For example, logicians and algebraists were strongly inspired by the problems in the database community, and database theoreticians learned relevant facts about semirings from the automata and algebra people. New collaborations emerged during the seminar, and there are some new papers with authors from different communities currently in the writing.
- We have assembled a list of open problems from all involved areas, which we included here. We hope that this list will help define the community interested in semirings, and will also inspire young researchers to contribute to this currently very timely area.
- A follow-up seminar is already being planned with a modified team of organizers (including two of the present ones, Guillermo Badia and Carles Noguera, together with Val Tannen and Thomas Eiter), adding the new community of AI.

Acknowledgements. We are grateful to the Scientific Directorate and to the staff of the Schloss Dagstuhl – Leibniz Center for Informatics for their support of this seminar. We also wish to express our sincere thanks to Sophie Brinke, Lovro Mrkonjić and Gaia Petreni for collecting the abstracts of the talks and compiling the list of open problems.

2 Table of Contents

3 Overview of Talks

3.1 The Semiring Framework for Data(base) Provenance

Val Tannen (University of Pennsylvania – Philadelphia, US)

Joint work of Yael Amsterdamer, Daniel Deutch, Nate Foster, Erich Grädel, Todd J. Green, Zachary G. Ives, Grigoris Karvounarakis, Tova Milo, Matthias Naaf, Sudeepa Roy, Val Tannen
Main reference Todd J. Green, Gregory Karvounarakis, Val Tannen: "Provenance semirings", in Proc. of the Twenty-Sixth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 11-13, 2007, Beijing, China, pp. 31–40, ACM, 2007.
URL https://doi.org/10.1145/1265530.1265535

Data from different sources mixes up in *data sharing* systems. We end up with complex relationships between participants' data. In well-disciplined systems, these relationships are specified by logical constraints that define *database transformations*. When participants receive data from the sharing system, they would like to apply some *trust* judgements to it. Their confidence may be based on external opinions about the other participants. To combine these opinions, they need to know how a data item they receive depends on the sources – other participants in the system. Knowing this means knowing the *provenance* of the data item.

This tutorial provides an overview of the *semiring framework* for databases and logics, which forms the foundation of the *semiring provenance* approach.

3.2 Weighted Automata and Weighted Logics

Erik Paul (Universität Leipzig, DE)

Weighted automata are a generalization of finite automata and allow the description of quantitative properties of languages. Finite automata themselves are known to be expressively equivalent to MSO logic due to the fundamental Büchi–Elgot–Trakhtenbrot Theorem. For a long time, a corresponding logical characterization of weighted languages recognizable by weighted automata had been missing.

In this tutorial, we provide an introduction to weighted automata and in particular to weighted logics, a generalization of classical MSO logic expressively equivalent to weighted automata. After giving some insights into the ideas behind weighted logics in their current form, we consider generalizations to data structures and automaton models beyond weighted automata over words as well as generalizations to weight structures beyond semirings. Finally, we present further evidence of the robustness of weighted logics by examining a generalization of the classical Feferman–Vaught Theorem to weighted logics over arbitrary finite structures and by showing how fragments of weighted logics capture languages recognizable by weighted automata restricted to a certain degree of ambiguity.

3.3 The Impact of State Mergingon Predictive Accuracy in Probabilistic Tree Automata: Dietze's Conjecture Revisited

Johanna Björklund (University of Umeå, SE)

Main reference Johanna Björklund: "The impact of state merging on predictive accuracy in probabilistic tree automata: Dietze's conjecture revisited", J. Comput. Syst. Sci., Vol. 146, p. 103563, 2024.
URL https://doi.org/10.1016/J.JCSS.2024.103563

Dietze's conjecture concerns the problem of equipping a tree automaton M with weights to make it probabilistic, in such a way that the resulting automaton N predicts a given corpus C as accurately as possible. The conjecture states that the accuracy cannot increase if the states in M are merged with respect to an equivalence relation $\sim$ so that the result is a smaller automaton $M^\sim$. Put differently, merging states can never improve predictions. This is under the assumption that both M and $M^\sim$ are bottom-up deterministic and accept every tree in C. We prove that the conjecture holds, using a construction that turns any probabilistic version $N^\sim$ of $M^\sim$ into a probabilistic version N of M, such that N assigns at least as great a weight to each tree in C as $N^\sim$ does.

3.4 Term Rewriting, Equality Saturation, the Chase, and Tree Automata

Dan Suciu (University of Washington – Seattle, US)

Joint work of Dan Suciu, Yisu Remy Wang, Yihong Zhang
Main reference Dan Suciu, Yisu Remy Wang, Yihong Zhang: "Semantic foundations of equality saturation", CoRR, Vol. abs/2501.02413, 2025.
URL https://doi.org/10.48550/ARXIV.2501.02413

Equality saturation is an approach to the word problem that is currently popular in the compilers community. In this talk we describe the equality saturation framework, and relate it to two well-known concepts: the chase of TGDs and EGDs, and to tree automata. We also describe open problems, some of which require extensions of the chase, or of tree automata, to semirings.

3.5 The Transportation Problem for Positive Commutative Monoids

Albert Atserias (UPC Barcelona Tech, ES)

Joint work of Albert Atserias, Phokion G. Kolaitis

In the transportation problem as studied in the theory of resource allocation in economic theory, a finite set of suppliers seek to distribute their production among a finite set of consumers to completely fill their demands. As an optimization problem, the goal is to achieve so at the minimum total transportation cost. Varying the interpretation of what supply, demand, and cost is, the transportation problem can be used to model a variety of tasks from different areas, from the original resource allocation problem in economics, to the problem of image reconstruction in tomography, the relational consistency problem in database theory, and the joint measurability problem of observables in quantum theory. Motivated by some of these applications, where the interpretations of supply and demand need not obey the arithmetic of the real numbers, we study the transportation problem in the abstract setting of positive semirings and, even more generally, in the setting of positive commutative monoids.

3.6 The Model Theory of Semiring Semantics

Erich Grädel (RWTH Aachen, DE)

Main reference Erich Grädel, Val Tannen: "Provenance Analysis and Semiring Semantics for First-Order Logic", CoRR, Vol. abs/2412.07986, 2024.
URL https://doi.org/10.48550/ARXIV.2412.07986

Semiring semantics relies on the idea to evaluate logical statements not just by true or false, but by values in some commutative semiring. This approach has originally been motivated by the provenance analysis of database queries, but has meanwhile been extended to a systematic alternative semantics for many logical systems, in particular for first-order logic and fixed-point logics. Semiring semantics can provide very detailed information about a query or a logical statement, for instance concerning the combinations of atomic facts that imply its truth, or practical information about evaluation costs, confidence scores, access levels, or the number of successful evaluation strategies.

The development of semiring semantics raises the question to what extent classical techniques and results of logic extend to semiring semantics, and how this depends on the algebraic properties of the underlying semiring. This tutorial provides a survey on recent results in this research programme, covering aspects such as axiomatisability of finite semiring interpretations, 0-1 laws, locality, and Ehrenfeucht–Fraïssé games.

3.7 The algebraic semantics of many-valued logics

Sara Ugolini (IIIA–CSIC – Bellaterra, ES)

Many-valued logics are systems of nonclassical logics which go beyond the true/false dichotomy of classical logic. By considering more than two truth values, they provide a formal framework for reasoning with partial truth and truth degrees.

In this tutorial we give an overview of many valued logics from the perspective of their algebraic semantics. In particular, we focus on those logics whose models can be presented by means of residuated structures and their reducts; these indeed cover most of the well-known many-valued logics considered in the literature, e.g., Łukasiewicz logic, Gödel–Dummett logic, all logics arising from (left-)continuous t-norms, 3-valued Kleene logic.

In particular, we discuss how algebraic methods can be fruitfully used to study logical properties.

References

1 N. Galatos, P. Jipsen, T. Kowalski, and H. Ono. *Residuated Lattices: An Algebraic Glimpse at Substructural Logics*, Studies in Logic and the Foundations of Mathematics, vol. 151, Elsevier, Amsterdam, The Netherlands, 2007.
2 P. Hájek. *Metamathematics of Fuzzy Logics*, Kluwer Academic Publisher, Dordrecht, The Netherlands, 1998.
3 G. Metcalfe, F. Paoli, and C. Tsinakis. *Residuated Structures in Algebra and Logic*, vol. 277. American Mathematical Society, 2023.

3.8 Compactness in Semiring Semantics

Sophie Brinke (RWTH Aachen, DE)

Joint work of Sophie Brinke, Anuj Dawar, Erich Grädel, Lovro Mrkonjić, Matthias Naaf

Semiring semantics for first-order logic evaluates formulae not just to true or false but to values from a commutative semiring. Depending on the underlying semiring, this allows us to track descriptions of the atomic facts that are responsible for the truth of a statement or practical information about the evaluation such as costs or confidence. Also classical semantics appears as a special case when the Boolean semiring is used, which raises the question to what extent model-theoretic results can be generalized beyond the Boolean semiring and how this relates to the algebraic properties of the underlying semiring. In this talk, we investigate the availability of compactness, which states that a set of sentences is satisfiable if every finite subset is. By defining satisfiability as the existence of non-zero valuations, this implication can be generalized to every absorptive semiring. For the, in Boolean semantics equivalent, formulation via entailment, the situation is different: Based on an order on the semiring, the entailment relation naturally extends to semiring semantics, but this yields a stronger variant of compactness, which fails for the tropical semiring, the Łukasiewicz semiring, and the semirings of generalized absorptive polynomials. However, it still holds for finite semirings and (possibly infinite) lattice semirings.

3.9 Codd's Theorem for Databases over Semirings

Guillermo Badia (University of Queensland – Brisbane, AU)

Joint work of Guillermo Badia, Phokion G. Kolaitis, Carles Noguera
Main reference Guillermo Badia, Phokion G. Kolaitis, Carles Noguera: "Codd's Theorem for Databases over Semirings", CoRR, Vol. abs/2501.16543, 2025.
URL https://doi.org/10.48550/ARXIV.2501.16543

Codd's Theorem, a fundamental result of database theory, asserts that relational algebra and relational calculus have the same expressive power on relational databases. We explore Codd's Theorem for databases over semirings and establish two different versions of this result for such databases: the first version involves the five basic operations of relational algebra, while in the second version the division operation is added to the five basic operations of relational algebra. In both versions, the difference operation of relations is given semantics using semirings with monus, while on the side of relational calculus a limited form of negation is used. The reason for considering these two different versions of Codd's theorem is that, unlike the case of ordinary relational databases, the division operation need not be expressible in terms of the five basic operations of relational algebra for databases over an arbitrary positive semiring; in fact, we show that this inexpressibility result holds even for bag databases.

3.10 Semirings with Infinitary Operations

Lovro Mrkonjić (RWTH Aachen, DE)

Joint work of Sophie Brinke, Erich Grädel, Lovro Mrkonjić, Matthias Naaf
Main reference Sophie Brinke, Erich Grädel, Lovro Mrkonjic, Matthias Naaf: "Semiring Provenance in the Infinite", in Proc. of the The Provenance of Elegance in Computation - Essays Dedicated to Val Tannen, Tannen's Festschrift, May 24-25, 2024, University of Pennsylvania, Philadelphia, PA, USA, OASIcs, Vol. 119, pp. 3:1–3:26, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024.
URL https://doi.org/10.4230/OASICS.TANNEN.3

Some applications of semirings require *infinitary semirings* with addition and multiplication operators over infinite collections of values. For example, in semiring semantics for first-order logic, existential and universal quantifiers over infinite domains are interpreted using infinitary addition and multiplication, respectively. In order to guarantee well-defined and informative semiring semantics in the infinitary setting, infinitary operations must satisfy core algebraic properties, such as distributivity and invariance under partition and re-ordering of the operands, which are analogous to associativity and commutativity from the finite setting.

We present an overview of infinitary semirings and precise definitions of their core algebraic properties, and we discuss further properties such as compactness. It turns out that some semirings do *not* admit infinitary operations, some semirings must be *extended* in order to allow infinitary operations, and other semirings, such as semirings induced by complete lattices, naturally admit infinitary operations. Finally, we outline the impact of infinitary operations on free polynomial semirings, which are used for provenance analysis in logic, databases and game theory. Our analysis shows that, with suitable definitions for infinitary semirings, large parts of the theory of semiring provenance can be successfully generalized to infinite domains.

3.11 Models of Computation over Semirings

Peter Kostolányi (Comenius University – Bratislava, SK)

Main reference Peter Kostolányi: "Weighted automata and logics meet computational complexity", Inf. Comput., Vol. 301, p. 105213, 2024.
URL https://doi.org/10.1016/J.IC.2024.105213

The tutorial highlights a recent reinterpretation of certain parts of computational complexity theory from the viewpoint of a weighted automata theorist. It explains how the framework of weighted Turing machines over semirings can be used to capture various sorts of quantitative complexity classes – such as those of counting and optimisation problems – in a consistent manner as instances of weighted complexity classes over different semirings. Several basic results about the classical complexity classes of decision problems turn out to generalise to this weighted setting, while new appealing connections to weighted logics arise. The tutorial gives a brief overview of the most substantial observations of this kind.

3.12 Hoops in semiring semantics

Tomasz Kowalski (Jagiellonian University – Kraków, PL)

Many semirings used in semiring semantics are naturally ordered and admit an additional subtraction-like operation, often called *monus*. It turns out that multiplication-free reducts of such semirings are very well known in universal algebra and algebraic logic under the name of *dual hoops*. A dual hoop is an algebra $\mathbf{A} = (A; +, -, 0)$ such that $(A; +, 0)$ is a commutative monoid, and moreover the following identities hold:

- $x - x = 0$,
- $(x - y) + y = (y - x) + x$,
- $x - (y + z) = (x - y) - z$.

This implies that $x \leq y \iff x - y = 0$ defines an order, $-$ is a co-residual of $+$ in this order, and moreover the order is natural. In algebraic logic, these algebras are more often presented in the order-dual form, with $\cdot$, $\rightarrow$ and 1, so from now on I will always speak of hoops, relying on the reader to identify the correct order.

Hoops were first considered by Bosbach in [1, 2], then studied by Büchi and Owens in [3], and then investigated thoroughly by Blok and Ferreirim in [4]. Hoops have a number of pleasant algebraic properties, and a number of them follow from something more general than the hoop structure, namely, from the fact that they are *subtractive*. The property of subtractivity was isolated and thoroughly studied by Aglianò and Ursini in a series of articles starting from [5] (with a prequel by Ursini). The hoop structure adds natural (in the technical sense) residuated order to subtractivity, yelding algebras resembling ordered rings.

Furthermore, expanding a hoop to a semiring preserves the pleasant algebraic properties alluded to above. The theory of such expansions was studied, somewhat more generally, by Blok and Pigozzi in [6].

My talk will be a sightseeing tour, introducing such expanded hoops and presenting a few examples: in particular showing that some commonly used semirings such as $\mathbb{N}[X]$ have an implicit dual hoop structure.

References

1. B. Bosbach. *Komplementäre Halbgruppen. Axiomatik und Arithmetik*, Fundamenta Mathematicae 64, 257–287, 1969.
2. B. Bosbach. *Komplementaere Halbgruppen. Kongruenzen und Quotienten*, Fundamenta Mathematicae 69, 1–14, 1970.
3. J. R. Büchi and T. M. Owens. *Complemented monoids and hoops*, unpublished manuscript.
4. W. J. Blok and I. M. A. Ferreirim. *Hoops and their implicational reducts*, Algebra Universalis 43, 233–257, 2000.
5. P. Aglianò and A. Ursini. *On subtractive varieties II: General properties*, Algebra Universalis 36, issue 2, 222–259, 1996.
6. W. J. Blok and D. Pigozzi. *On the structure of varieties with equationally definable principal congruences III*, Algebra Universalis 32, 545–608, 1994.

3.13 Valued Constraint Satisfaction Problem and Resilience in Database Theory

Žaneta Semanišinová (TU Dresden, DE)

Joint work of Manuel Bodirsky, Carsten Lutz, Žaneta Semanišinová
Main reference Manuel Bodirsky, Zaneta Semanisinová, Carsten Lutz: "The Complexity of Resilience Problems via Valued Constraint Satisfaction Problems", in Proc. of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2024, Tallinn, Estonia, July 8-11, 2024, pp. 14:1–14:14, ACM, 2024.
URL https://doi.org/10.1145/3661814.3662071

A recent research topic in database theory is the computational complexity of resilience of queries. For a fixed conjunctive query, the problem is to compute the number of facts that need to be removed from a given database so that it does not satisfy the query. Mathematically, this problem can be viewed as removing tuples from relations of a relational structure so that it does not satisfy a fixed primitive positive sentence. In this talk, I will explain how resilience problems can be viewed as valued constraint satisfaction problems (VCSPs) of structures that are finite or at least finite-like (in the sense that they have an oligomorphic automorphism group). A VCSP is parameterized by a valued structure, which consists of a domain and cost functions, each defined on some finite power of the domain with values in the semiring of rational numbers extended with positive infinity. Building on the known results about VCSPs on finite domains, we explore how the setting can be generalized to infinite domains and give some general powerful hardness and tractability conditions for the resilience problem.

3.14 Polynomial-time Convergence of Datalogo over p-stable Semirings

Hung Ngo (relationalAI – Berkeley, US)

Joint work of Sungjin Im, Ben Moseley, Hung Ngo, Kirk Pruhs
Main reference Sungjin Im, Benjamin Moseley, Hung Q. Ngo, Kirk Pruhs: "Polynomial Time Convergence of the Iterative Evaluation of Datalogo Programs", CoRR, Vol. abs/2312.14063, 2023.
URL https://doi.org/10.48550/ARXIV.2312.14063

Datalogo is an extension of Datalog that allows for aggregation and recursion over an arbitrary commutative semiring. Unlike Datalog, the natural iterative evaluation of some Datalogo programs over some semirings may not converge. It is known that the commutative semirings for which the iterative evaluation of Datalogo programs is guaranteed to converge are exactly those semirings that are stable.

Previously, the best known upper bound on the number of iterations until convergence over p-stable semirings is exponential in data complexity. In this talk, we present a recent finding that the native iterative algorithm for Datalogo converges in polynomial time.

3.15 Not all semirings are strong

Jacques Sakarovitch (IRIF, CNRS – Université Paris Cité, FR & LTCI, Télécom Paris, IPP, FR)

Main reference Jacques Sakarovitch: "Elements of Automata Theory". Cambridge University Press, 2009.
URL https://doi.org/10.1017/CBO9781139195218

Working with weighted automata naturally leads to the problem of the definition of the (Kleene) star operation on the semiring of (formal power) series. And proving that rational series are realised by weighted finite automata (one direction of the Kleene Theorem) may require to establish the identity: $(s_0 + s_p)^* = (s_0)^*(s_p(s_0)^*)^*$ where s_0 and s_p are respectively the "constant term" and the "proper part" of a series s.

In my book "Elements of Automata Theory", I gave a proof of the above identity under the hypothesis that the weight semiring has the property that the product of two summable families is a summable family. I called such semirings "strong" and even though all semirings that I knew are strong, I stated the conjecture that there should exist some semirings which are not strong.

I have presented the problem but for want of time, I have not been able to describe the example of a semiring which is not strong that my colleague David Madore gave me and that will be the subject of a forthcoming publication. At least, I was able to announce the existence of such a semiring which justifies the title.

3.16 Semiring Circuits for Sum-Product Queries

Paris Koutris (University of Wisconsin-Madison, US)

Joint work of Austen Fan, Paris Koutris, Hangdong Zhao
Main reference Austen Z. Fan, Paraschos Koutris, Hangdong Zhao: "Tight Bounds of Circuits for Sum-Product Queries", Proc. ACM Manag. Data, Vol. 2(2), p. 87, 2024.
URL https://doi.org/10.1145/3651588

In this talk, we ask the following question: given a Boolean Conjunctive Query, what is the smallest circuit that computes the provenance polynomial of the query over a given semiring? We answer this question by giving upper and lower bounds. We show a circuit construction that matches this bound when the semiring is idempotent. The techniques we use combine several central notions in database theory: provenance polynomials, tree decompositions, and disjunctive Datalog programs. We extend our results to lower and upper bounds for formulas. Finally, we briefly present open problems related to circuit constructions.

3.17 Rewriting Consistent Answers on Annotated Data and Semiring Circuits

Jonni Virtema (University of Sheffield, GB)

Joint work of Phokion G. Kolaitis, Nina Pardal, Jonni Virtema
Main reference Phokion G. Kolaitis, Nina Pardal, Jonni Virtema: "Rewriting Consistent Answers on Annotated Data", CoRR, Vol. abs/2412.11661, 2024.
URL https://doi.org/10.48550/ARXIV.2412.11661

We study consistent answers of queries over databases annotated with values from a naturally ordered positive semiring. In this setting, the consistent answers of a query are defined as the minimum of the semiring values that the query takes over all repairs of an inconsistent database. The main focus is on self-join free conjunctive queries and key constraints, which is the most extensively studied case of consistent query answering over standard databases. We introduce a variant of first-order logic with a limited form of negation, define suitable semiring semantics, and establish that the consistent query answers of a self-join free conjunctive query under key constraints are rewritable in this logic if and only if the attack graph of the query contains no cycles. This result generalizes an analogous result of Koutris and Wijsen for ordinary databases, but also yields new results for a multitude of semirings, including the bag semiring, the tropical semiring, and the fuzzy semiring. Moreover, we relate our logic to particular arithmetic circuit complexity classes defined over semirings.

3.18 When do homomorphism counts help in query algorithms?

Balder ten Cate (University of Amsterdam, NL)

Joint work of Balder ten Cate, Victor Dalmau, Phokion G. Kolaitis, Wei-Lin Wu
Main reference Balder ten Cate, Victor Dalmau, Phokion G. Kolaitis, Wei-Lin Wu: "When Do Homomorphism Counts Help in Query Algorithms?", in Proc. of the 27th International Conference on Database Theory, ICDT 2024, March 25-28, 2024, Paestum, Italy, LIPIcs, Vol. 290, pp. 8:1–8:20, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024.
URL https://doi.org/10.4230/LIPICS.ICDT.2024.8

A query algorithm based on homomorphism counts is a procedure for determining whether a given instance satisfies a property by counting homomorphisms between the given instance and finitely many predetermined instances. In a left query algorithm, we count homomorphisms from the predetermined instances to the given instance, while in a right query algorithm we count homomorphisms from the given instance to the predetermined instances. Homomorphisms are usually counted over the semiring $\mathbb{N}$ of non-negative integers; it is also meaningful, however, to count homomorphisms over different semirings, including the Boolean semiring $\mathbb{B}$, where the homomorphism count indicates whether or not a homomorphism exists. I will present some results from joint work with Victor Dalmau, Phokion G. Kolaitis and Wei-Lin Wu (ICDT 2024), where we compare the relative expressive power of query algorithms over $\mathbb{N}$ and over $\mathbb{B}$. One of the main results asserts that if a property is closed under homomorphic equivalence, then that property admits a left query algorithm over $\mathbb{B}$ if and only if it admits a left query algorithm over $\mathbb{N}$.

3.19 How to Bake an Uncertainty Pie: Take Even Parts of Abstract Interpretation, K-relations, and Zonotopes and Mix Thoroughly

Boris Glavic (University of Illinois – Chicago, US)

Joint work of Su Feng, Boris Glavic, Oliver Kennedy, Babak Salimi
Main reference Jiongli Zhu, Su Feng, Boris Glavic, Babak Salimi: "Learning from Uncertain Data: From Possible Worlds to Possible Models", in Proc. of the Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024, 2024.
URL http://papers.nips.cc/paper_files/paper/2024/hash/c17fab1bcef325d0d30989c9bf506c0b-Abstract-Conference.html

Uncertainty arises naturally due to data quality issues such as missing values, constraint violations, and outliers. As the ground truth clean version of a dirty database is typically not recoverable, there is a need to reason about all possible outcomes of computations over all possible clean versions of the database and to determine which outcomes will be certainly returned by the computation irrespective of uncertainty about the ground truth clean dataset. These problems have been studied in reachability analysis in control theory and in the context of certain answers and consistent query answering in databases. As both problems are known to be computationally hard for many interesting classes of computation, there is a need for conservative approximations: under-approximate what is certainly known to be true and over-approximate what is possibly true. In this talk, I will introduce techniques for computing such approximations for relational queries, yielding a query semantics for uncertain data for full relational algebra queries with aggregation and sorting-based operations that has PTIME data complexity and an approach for training and inference with linear models where both the training data and test data may be dirty. On the side of over-approximating all possible outcomes both approaches are shown to be instances of abstract interpretation, a technique from programming languages for over-approximating a set of inputs as an element from a so-called abstract domain and for computations in the abstract domain that preserve this over-approximation. In the cases of relational queries, K-relations over a naturally-ordered semiring K can be used as the abstract domain and standard K-relational query semantics is sufficient to preserve the over-approximation. To deal with aggregation and value uncertainty, the domain of these K-relations are intervals. In case of machine learning, the abstract domain of zonotopes, a restricted type of convex polytopes expressed as affine transformations of the unit hypercube, is used to encode uncertainty. This domain has the advantage of being able to encode correlations between values, but necessitates new methods for computing closed form solutions to models that will be briefly introduced in this talk.

References

1 Su Feng, Boris Glavic, Oliver Kennedy: *Efficient Approximation of Certain and Possible Answers for Ranking and Window Queries over Uncertain Data*, Proceedings of the VLDB Endowment 16(6), 1346–1358, 2023.
2 Su Feng, Boris Glavic, Aaron Huber, Oliver Kennedy: *Efficient Uncertainty Tracking for Complex Queries with Attribute-Level Bounds*, Proceedings of the 46th International Conference on Management of Data, 528–540, 2021.
3 Su Feng, Aaron Huber, Boris Glavic, Oliver Kennedy: *Uncertainty Annotated Databases - a Lightweight Approach for Approximating Certain Answers*, Proceedings of the 44th International Conference on Management of Data, 1313–1330, 2019.

3.20 Complete Additively Idempotent Semirings and their Applications in Weighted Automata Theory, Social Network Analysis and Modal Logics

Miroslav Ćirić (University of Niš – Serbia, RS)

Joint work of Miroslav Ćirić, Jelena Ignjatović, vana Micić, Stefan Stanimirović, Aleksandar Stamenković, Nada Damljanović, Zorana Jančić, Ivan Stanković, Jelena Matejić

An additively-idempotent semiring is a semiring whose addition is an idempotent operation. In such a semiring, one can define a partial ordering, called the natural partial ordering, which is compatible both with addition and multiplication, and has the zero as the smallest element. Relative to the natural partial ordering, an additively-idempotent semiring is positively ordered, and besides, it forms an upper semilattice in which the binary supremum operation coincides with addition. In the case when that upper semilattice is complete, in the sense that every subset, whether finite or infinite, has a supremum, then it is convenient to treat infinite suprema as infinite sums, seeing that they satisfy all properties from the definition of infinite addition. If, in addition, infinite distributive laws hold, then we say that the considered semiring is a complete additively-idempotent semiring. The class of such semirings is quite broad and includes many important types of semirings, such as complete max-plus semirings, complete min-plus (tropical) semirings, as well as various semirings that are semiring reducts of algebras widely used in fuzzy set theory.

The key property of complete additively-idempotent semirings is the existence of residuation, that is, the existence of a pair of binary operations adjoined to multiplication, which enable certain types of inequations and equations to be solved in such a semiring. In addition, residuation is transferred to matrices and vectors over such a semiring, which gives the possibility to solve various inequalities and equations involving matrices and vectors. The purpose of this talk is to show how residuation can be applied in solving some very important problems in weighted automata theory (containment and equivalence, simulations and bisimulations, state reduction), in social network analysis (positional analysis, blockmodeling) and modal logics.

3.21 On the Boolean Closure of Deterministic Tree Automata

Wolfgang Thomas (RWTH Aachen, DE)

Joint work of Christof Löding, Wolfgang Thomas
Main reference Christof Löding, Wolfgang Thomas: "On the Boolean Closure of Deterministic Top-Down Tree Automata", Int. J. Found. Comput. Sci., Vol. 35(1&2), pp. 11–22, 2024.
URL https://doi.org/10.1142/S0129054123480015

The tree languages recognizable by deterministic top-down tree automata form a proper subclass D of the class of regular tree languages, and this proper containment also applies to its (strictly larger) Boolean closure $\mathrm{BC}(D)$. We present two results around the decades-long open problem whether for regular tree languages one can decide membership in $\mathrm{BC}(D)$. First, a characterization of $\mathrm{BC}(D)$ by a natural extension of deterministic top-down tree automata is presented, giving a convenient method to show that certain regular tree languages are outside $\mathrm{BC}(D)$. Secondly, it is shown that, for fixed k, it is decidable whether a regular tree language is a Boolean combination of k tree languages of D.

4 Open problems

4.1 Circuit Size for Reachability

Paris Koutris (University of Wisconsin-Madison, US)

Let K_n be the complete graph over n vertices. Consider the provenance polynomial that sums over all simple paths from node 1 to node n in K_n, and for every path takes the product of the weights x_{ij} of every edge $\{i, j\}$ in the path. We interpret sum and product over an absorptive semiring (absorptive means that $1 + x = 1$ for every element x).

The question is: What is the size of the smallest circuit that represents this polynomial? We know that we can construct such a circuit of size $\mathcal{O}(n^3)$ for any absorptive semiring [1]. But we do not know whether this is optimal, or whether we can show a matching lower bound. The lower bound is open even for the case of the tropical semiring. The best lower bound is the trivial $\Omega(n^2)$ bound that says that we need to see all input edges.

References

1 Stasys Jukna. *Lower Bounds for Tropical Circuits and Dynamic Programs*, Theory Comput. Syst. 57(1): 160–194, 2015. DOI: 10.1007/s00224-014-9574-4

4.2 Convergence rate of Datalogo over p-stable semirings

Hung Ngo (relationalAI – Berkeley, US)

Main reference Sungjin Im, Benjamin Moseley, Hung Q. Ngo, Kirk Pruhs: "Polynomial Time Convergence of the Iterative Evaluation of Datalogo Programs", CoRR, Vol. abs/2312.14063, 2023.
URL https://doi.org/10.48550/ARXIV.2312.14063

In the paper "Polynomial Time Convergence of the Iterative Evaluation of Datalogo Programs" to be presented at PODS 2025, we proved that a Datalogo program converges after a polynomial number of steps. The convergence time is roughly $\mathcal{O}(pn^4)$, where n is the output size. The best known lower bound is $\Omega((p + 1)n)$.

The open problem is to close this gap. We think that an upper bound of $\mathcal{O}((p + 1)n)$ is possible.

4.3 Subreducts of monus semirings

Paolo Aglianò (University of Siena, IT)

Many semirings have a natural structure of "semiring with monus", i.e. they possess a term operation which acts as a dual implication and it forms a residuated pair with + w.r.t. the natural ordering. However always asking such term to be present can cause unwanted problems with the current understanding of the theory. Of course it is possible to formulate the definitions in such a way that a semiring with monus is an algebra (in the universal algebraic sense). This suggests that a way out could be solving the following problem.

▶ **Problem.** Characterize those semirings that are subreducts of semirings with monus. Alternatively, characterize those semirings that are embeddable in a semiring with monus.

This could be useful because

1. from the general theory we know that this class is at least a quasivariety, so it has strong structural properties;
2. the embedding allows one to use the monus operation "as if it were present", but without having it interfere with the (quasi)equational properties of the class.

4.4 Weighted First-Order Definable Languages

Manfred Droste (Universität Leipzig, DE)

Joint work of Manfred Droste, Paul Gastin

▶ **Theorem.** *For $L \subseteq \Sigma^*$, the following statements are equivalent.*

1. *L is star-free.*
2. *L is aperiodic.*
3. *L is* FO*-definable.*
4. *L is* LTL*-definable.*

The equivalences of statements 1 and 2, 2 and 3 respectively 3 and 4 are fundamental theorems of Schützenberger [1], McNaughton and Papert [2] resp. Kamp [3]. A proof of these and more equivalences can be found in the survey by Diekert and Gastin [4]. This raises the question whether analogous results hold for weighted languages $s\colon \Sigma^* \to \mathcal{S}$, where $\mathcal{S}$ is a semiring.

$1'$ It is still unclear how statement 1 shall be translated to the weighted setting. A star-free language in the classical sense is a language that can be constructed from finite languages using Boolean operators, including complementation, and concatenation, but no Kleene star. In the semiring setting, the problem arises what to do with the complement.

$2'$ For statement 2, an adequate translation is already known: s is called aperiodic if $s = [\![\mathfrak{A}]\!]$ for some aperiodic, polynomially ambiguous NFA $\mathfrak{A}$. Polynomially ambiguous means that $\mathfrak{A}$ shall only admit polynomially many paths for any input.

$3'$ Statement (3) can be translated to: s is definable in weighted FO.

$4'$ It is still open to define a suitable weighted version of LTL.

First results on formulations of statements $1'$ and $2'$ and their equivalence are contained in [5]. An equivalence of statements $2'$ and $3'$ was given more recently in [6].

There are natural formulations of a weighted LTL for which the implication $4'$ implies $3'$ is straightforward. The difficulty is to obtain an equivalence. An ideal result would be a formulation of a weighted LTL which, possibly through the implication of $4'$ implies $2'$, would lead to procedures for quantitative model checking, similar to the seminal developments for the unweighted case.

In summary, it is still open to find adequate formulations for statements $1'$ and $4'$ and to prove or disprove their equivalence to the remaining statements.

References

1 M.-P. Schützenberger. *On Finite Monoids Having Only Trivial Subgroups*, Information and Control 8(2): 190–194, 1965. DOI: 10.1016/s0019-9958(65)90108-7
2 R. McNaughton and S. Papert. *Counter-free Automata*, Research Monograph, vol. 65, MIT Press, 1971.
3 J. A. W. Kamp. *Tense Logic and the Theory of Linear Order*, PhD Thesis, University of California, Los Angeles (UCLA), 1968.
4 V. Diekert and P. Gastin. *First-order definable languages*, Logic and Automata: History and Perspectives (J. Flum, E. Grädel, Th. Wilke, eds.), Amsterdam University Press, 2008, pp. 261–306. ISBN: 978-90-5356-576-6.
5 M. Droste and P. Gastin. *On aperiodic and star-free formal power series in partially commuting variables*, Formal Power Series and Algebraic Combinatorics (D. Krob, A. A. Mikhalev, A. V. Mikhalev, eds.), 12th Int. Conf., Moscow, Springer, 2000, pp. 158–169.
6 M. Droste and P. Gastin. *Aperiodic weighted automata and weighted first-order logic*, Mathematical Foundations of Computer Science (MFCS 2019), LIPIcs, vol. 138, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019, pp. 76:1–76:15.

4.5 Conjunctive Query Containment under Bag Semantics

Phokion G. Kolaitis (University of California – Santa Cruz, US & IBM Almaden Research Center – San Jose, US)

The conjunctive query containment problem asks: given two conjunctive queries q and q' of the same arities, is it true that on every relational database D, it holds that $q(D) \subseteq q'(D)$?

In 1977, Chandra and Merlin showed that, for standard relational databases, this problem is NP-complete [1]. In 1993, Chaudhuri and Vardi raised the question of whether or not the conjunctive query containment problem is decidable for bag databases, i.e., when the databases considered are databases over the semiring $(\mathbb{N}, +, \times, 0, 1)$ of the natural numbers [2]. In spite of concerted efforts by several different groups of researchers, the conjunctive query containment problem for bag databases remains open to date. It is known, however, that it the containment problem for bag databases becomes undecidable if slightly larger classes of queries are considered, such as unions of conjunctive queries [3] or conjunctive queries that include negated equalities as atoms [4].

It should be noted that the conjunctive query containment problem is also meaningful for databases over naturally ordered semirings. Results concerning the decidability and complexity of this problem over such semirings have been obtained by T. J. Green in 2011 [5] and by Kostylev, Reutter, and Salamon in 2014 [6].

References

1 A. K. Chandra and P. M. Merlin. *Optimal implementation of conjunctive queries in relational data bases*, STOC 1977: 77–90. DOI: 10.1145/800105.803397
2 S. Chaudhuri and M. Y. Vardi. *Optimization of Real Conjunctive Queries*, PODS 1993: 59–70. DOI: 10.1145/153850.153856
3 Y. E. Ioannidis and R. Ramakrishnan. *Containment of Conjunctive Queries: Beyond Relations as Sets*, ACM Trans. Database Syst. 20(3): 288–324, 1995. DOI: 10.1145/211414.211419
4 T. S. Jayram, P. G. Kolaitis, E. Vee. *The Containment Problem for Real Conjunctive Queries with Inequalities*, PODS 2006: 80–89. DOI: 10.1145/1142351.1142363

5 T. J. Green. *Containment of Conjunctive Queries on Annotated Relations*, Theory Comput. Syst. 49(2): 429–459, 2011. DOI: 10.1007/s00224-011-9327-6

6 E. V. Kostylev, J. L. Reutter, A. Z. Salamon. *Classification of Annotation Semirings over Containment of Conjunctive Queries*, ACM Trans. Database Syst. 39(1): 1:1–1:39, 2014. DOI: 10.1145/2556524

Participants

- Samson Abramsky
University College London, GB
- Paolo Aglianò
University of Siena, IT
- Antoine Amarilli
Inria Lille – Nord Europe, FR
- Albert Atserias
UPC Barcelona Tech, ES
- Guillermo Badia
University of Queensland – Brisbane, AU
- Johanna Björklund
University of Umeå, SE
- Camille Bourgaux
CNRS & ENS – Paris, FR
- Sophie Brinke
RWTH Aachen, DE
- Silvia Butti
University of Oxford, GB
- Miroslav Ćirić
University of Niš – Serbia, RS
- Víctor Dalmau
UPF – Barcelona, ES
- Frank Drewes
University of Umeå, SE
- Manfred Droste
Universität Leipzig, DE
- Thomas Eiter
TU Wien, AT
- Boris Glavic
University of Illinois – Chicago, US
- Erich Grädel
RWTH Aachen, DE
- Marcel Jackson
La Trobe University – Bundoora, AU
- Phokion G. Kolaitis
University of California – Santa Cruz, US & IBM Almaden Research Center – San Jose, US
- Peter Kostolányi
Comenius University – Bratislava, SK
- Egor Kostylev
University of Oslo, NO
- Paris Koutris
University of Wisconsin-Madison, US
- Tomasz Kowalski
Jagiellonian University – Kraków, PL
- Dietrich Kuske
TU Ilmenau, DE
- Filip Mazowiecki
University of Warsaw, PL
- Lovro Mrkonjić
RWTH Aachen, DE
- Hung Ngo
relationalAI – Berkeley, US
- Carles Noguera
University of Siena, IT
- Nina Pardal
University of Huddersfield, GB
- Erik Paul
Universität Leipzig, DE
- Vitaly Perevoshchikov
Herrischried, DE
- Gaia Petreni
University of Siena, IT
- Jacques Sakarovitch
IRIF, CNRS – Université Paris Cité, FR & LTCI, Télécom Paris, IPP, FR
- Žaneta Semanišinová
TU Dresden, DE
- Dan Suciu
University of Washington – Seattle, US
- Val Tannen
University of Pennsylvania – Philadelphia, US
- Balder ten Cate
University of Amsterdam, NL
- Wolfgang Thomas
RWTH Aachen, DE
- Sara Ugolini
IIIA–CSIC – Bellaterra, ES
- Jonni Virtema
University of Sheffield, GB

Report from Dagstuhl Seminar 25082

Visualizing Data on Non-Flat, Non-Rectangular Displays

Anastasia Bezerianos*[1], Raimund Dachselt*[2], Wesley J. Willett*[3], and Ricardo Langner†[4]

1 **Université Paris-Saclay, INRIA, CNRS, FR.** anastasia.bezerianos@universite-paris-saclay.fr
2 **Technische Universität Dresden, DE.** raimund.dachselt@tu-dresden.de
3 **University of Calgary, CA.** wj@wjwillett.net
4 **Technische Universität Dresden, DE.** ricardo.langner@tu-dresden.de

Abstract

Experiences with data, visualization, and computing more broadly are mediated via flat, rectangular displays. However, an exciting range of new display technologies, including flexible, spherical, physical, and even robotic and drone-based displays, have recently emerged and are increasingly commercially available. These novel types of displays offer new ways to represent, explore, communicate, and share data, yet there is very little understanding of how to best utilize these novel form factors for data visualizations. This Dagstuhl Seminar aimed to escape from the display flatland that characterizes research in visualization, and to create a roadmap for future research on interactive non-flat displays. Bringing together researchers from data visualization, human-computer interaction, ubiquitous computing, tangible interaction, mobile and wearable technologies, and design, we surveyed the landscape of emerging technologies, ideated future opportunities for visualization on non-flat displays, and outlined a common research agenda for this emerging area.

Seminar February 16–21, 2025 – https://www.dagstuhl.de/25082
2012 ACM Subject Classification Human-centered computing → Human computer interaction (HCI); Human-centered computing → Visualization; Hardware → Emerging technologies
Keywords and phrases data visualization, design and evaluation, human-computer interaction, non-flat non-rectangular displays, visualizations beyond the desktop
Digital Object Identifier 10.4230/DagRep.15.2.110

1 Executive Summary

Anastasia Bezerianos (Université Paris-Saclay, INRIA, CNRS, FR)
Raimund Dachselt (Technische Universität Dresden, DE)
Wesley J. Willett (University of Calgary, CA)

The physical screens that we consider when creating visualizations for representing data, making decisions with this data, and learning about our environment – mobile phones and tablets, desktops, or even large wall or table displays – are predominantly flat and rectangular. Nevertheless, technology companies are actively pursuing a wide range of new display technologies, including curved displays, bendable displays, spherical displays, light field displays, cube displays, physical pin-based displays, foldable phones, and even robotic

* Editor / Organizer
† Editorial Assistant / Collector

and drone displays. Yet to date, visualization researchers and practitioners have not deeply considered the true potential for these new form factors, nor their challenges when it comes to communicating data.

This Dagstuhl Seminar brought together a diverse set of visualization and human-computer interaction researchers to outline an agenda for visualization on interactive non-flat displays and prepare the visualization research community for a future in which data displays become richer, more interactive, and more deeply embedded in environments. To that end, the seminar paired state-of-the-art talks and show-and-tell activities designed to build familiarity with a wider range of novel display technologies with ideation and design futuring sessions focused on envisioning their visualization potential.

Initial presentations included overviews of a range of new enabling technologies, including light field displays, shape-changing surfaces, and robotic displays, along with late-breaking work in situated visualization and data physicalization. Show-and-tell sessions also featured a diversity of physical and digital display types, with participants demonstrating novel research prototypes and commercial displays that push the limits of current display technologies. To push our vision of possible non-flat visualizations further, seminar participants then engaged in a series of design futuring activities, using both traditional sketching and generative AI tools to envision and elaborate possible future displays, visualizations, and scenarios. Building upon these generative activities, participants then worked together to identify and connect key research opportunities and challenges for visualization on non-flat displays, laying the groundwork for a research agenda for the field and also identifying a variety of targeted follow-up projects.

2 Table of Contents

3 Structure of the Seminar

3.1 Week at a Glance

Anastasia Bezerianos (Université Paris-Saclay, INRIA, CNRS, FR)
Raimund Dachselt (Technische Universität Dresden, DE)
Wesley J. Willett (University of Calgary, CA)

Monday

The seminar was opened by the organizers with an introduction to the topic of data visualization on non-flat non-rectangular displays (NFNRD), by presenting objectives and potential outcomes, and by providing organizational information. This was followed by a fast-paced self-introduction of all participants, where each person was giving a lightning talk about their background, relation to the seminar's topic, and expectations from the seminar.

An important activity to connect people and to foster collaboration was the research speed-dating conducted before and after lunch in two rounds. One half of the seminar participants was seated, while the half rotated in five minute time-slots to quickly brainstorm and discuss potential joint research. This activity helped bring together people in a fast-paced and focused fashion while also seeding follow-up paper activities.

Altogether, we scheduled five invited overview talks at the first three days to introduce a particular related topic area from the perspective of one expert participant. In the afternoon, Emmanuel Pietriga started with the first talk on the topic "An overview of (Wearable) Light Field Displays" (see Sec. 4.1).

The final scheduled activity of the day was a group activity elicitating the state-of-the-art of visualization beyond flat, rectangular display from four different topic lenses:

- Interaction and HCI (see Sec. 5.1)
- Perception, Design, and Visualization (see Sec. 5.2)
- Data Physicalization & Shape-Changing Displays (see Sec. 5.3)
- AR/MR/XR Visualization (see Sec. 5.4)

Tuesday

The second day of the seminar was started with the second overview talk given by Nathalie Bressa on the topic "Situated Visualization & Physicalization" (see Sec. 4.2). Afterwards, all state-of-the-art groups from the previous day reported back on their findings and informed all other seminar participants about the key research contributions, advances, publications, research venues, and community activities in the respective area.

An important activity central to the seminar's core objective of creating innovative designs and envisioning novel visualization approaches was the show-and-tell session conducted in two rounds before lunch. Many participants brought their own demos, early research prototypes, videos, material probes, and sketches to let others experience and discuss inspiring research ideas (see Fig. 1).

After lunch, Wesley Willett gave the third overview talk on "Visualization Futures" (see Sec. 4.3). This was intended as a starting point for the following afternoon group activities, two design futuring sessions. Here, participants split into six smaller groups and used the card-based sketching game "Vis Futures"[1], where players think critically (and playfully)

[1] https://dataexperience.cpsc.ucalgary.ca/visfutures/

Figure 1 The Tuesday show-and-tell session allowing participates to showcase, experience, and discuss latest research prototypes, materials, and ideas.

about the future of data and visualization. Using sketches, AI-generated images, videos, and small physical prototypes, a series of innovative design ideas was developed and presented in a plenary to all seminar participants. Groups explored diverse concepts ranging from visualization on deformable shape-changing displays to vegetable bio-displays, textile and garment displays, foldable displays, and water displays.

Wednesday

The first part of the morning was spent on two more overview talks concluding the series. First, Aluna Everitt talked about “Democratizing the Design and Development of Emerging Technologies” (see Sec. 4.4), followed by a surveying talk on the topic of “AR x Human-Robot Interaction” given by Ryo Suzuki (see Sec. 4.5). Of course, we also followed the great tradition of a group picture on the castle’s staircase (see p. 125).

To generate breakout groups for discussing the grand challenges of visualization for NFNRD, all attendees spent the remaining time before lunch brainstormed critical topics, challenges, and technologies. We then used affinity diagramming and clustering to consolidate these ideas (see Fig. 2, middle, for some of the resulting post-it groups).

After assigning people to six breakout groups based on the emergent themes from the clustering exercise, we enjoyed lunch followed by a social activity with many research-related and personal discussions, a bus trip to the city of Trier. There, we participated in a guided tour of the historic city and had a Roman-inspired dinner at the restaurant “Zum Domstein”, which enabled new bonds to be forged between the participants.

Thursday

This day saw intense and deep scientific discussions across several breakout group work sessions. Spread over the whole day and with reporting-back sessions in between (see, for example, Fig. 2), the discussion centered around the following topics:

- What is the scope of visualization on NFNRD? (11 participants, see Sec. 6.1)
- Display and input technologies (6 participants, see Sec. 6.2)
- Applications: What is the benefit of visualizing data on NFNRD? (6 part., see Sec. 6.3)
- Relationship between data and displays (7 participants, see Sec. 6.4)
- Designing visualizations for NFNRD (6 participants, see Sec. 6.5)
- Evaluating visualizations on NFNRD (5 participants, see Sec. 6.6)

Figure 2 People from the scoping breakout group reporting back to all participants.

Some groups discussed the same topics across several sessions, but with shifting group members. The summaries provided in Sec. 6 provide an aggregated overview of the major findings that, taken together, lay the foundations for a research agenda with challenges and opportunities.

Friday

The entire morning was spent wrapping up the seminar. First, six concrete research projects and publication ideas that emerged during the past days were openly discussed among participants, along with potential publication venues, such as IEEE VIS, ACM CHI, IEEE TVCG, and others. Second, a bigger joint publication on the challenges of data visualization on non-flat, non-rectangular displays was discussed and sketched in its structure. We agreed on further steps and publication venues. Especially with this second effort, we aim to outline a comprehensive research agenda for data visualization on these emerging display types. This shall enable other researchers and practitioners in the fields of visualization, HCI and visual design to develop future data visualization solutions for these new form factors – thereby going beyond display flatland.

4 Overview of Talks

4.1 An Overview of Light Field Displays

Emmanuel Pietriga (INRIA Saclay – Orsay, FR)

I presented an overview of light field displays by first introducing light field photography. Light field displays enable virtual content to be rendered at different focal depths simultaneously and are particularly interesting for Augmented Reality. They enable blending virtual content with the physical world, as the light from virtual elements behaves similar to the light from the physical environment. It also has the potential to solve the vergence accommodation conflict. Virtual content is not limited to a single focal plane and can also be rendered much closer to the user's eyes. Light field displays are still in their infancy, and I talked about the visual perception studies we are running to better understand their potential benefits.

4.2 Situated Visualization & Physicalization

Nathalie Bressa (Télécom Paris, FR)

Situated visualization is an area within information visualization that focuses on placing data visualizations in physical places to display contextually relevant information that can be related to people, places, or objects. Physicalizations, on the other hand, are physical artifacts that encode data through their geometry or material properties. In this talk, I provided an overview of these two concepts and highlighted their connections and overlap. By discussing a diverse set of examples of situated visualizations and physicalizations, I aimed to inspire new ideas for visualization on non-flat, non-rectangular displays.

4.3 Visualization Futures

Wesley J. Willett (University of Calgary, CA)

Together, let's consider visualization futures — imagining possible and probable future technologies, probing their implications, and using them as inspiration for visualization research. This talk highlights the potential for design futuring to (1) draw ideas from existing science fiction and speculative media, (2) craft new future visions that explore, provoke, and critique, and (3) articulate new trajectories, frameworks, and theories for data visualization as a field. Using the Vis Futures card deck, plus a range of generative and manual design tools, we'll also ideate and elaborate new visualizations for a variety of novel display platforms.

4.4 Democratizing the Design and Development of Emerging Technologies

Aluna Everitt (University of Canterbury – Christchurch, NZ)

My research focuses on democratizing the development of emerging technologies. More specifically, by establishing accessible approaches for designing and building emerging technologies such as robotics, wearables, and shape-changing interfaces. To advance the field, my research focuses not only on understanding these technologies (e.g., their design), but also how to build them (e.g., engineer them), and how to innovate with them (e.g., application). In this talk, I will go over some of the projects I have worked on around this topic across the fields of HCI, Design, and Engineering.

4.5 AR × Human-Robot Interaction

Ryo Suzuki (University of Colorado – Boulder, US & Tohoku University – Sendai, JP)

In this talk, I gave an overview of research examples and a taxonomy of augmented reality and robotics based on my CHI 2022 paper. Augmented and mixed reality (AR/MR) have emerged as a new way to enhance human-robot interaction (HRI) and robotic interfaces (e.g., actuated and shape-changing interfaces). Recently, an increasing number of studies in HCI, HRI, and robotics have demonstrated how AR enables better interactions between people and robots. However, often research remains focused on individual explorations and key design strategies, and research questions are rarely analyzed systematically. In this paper, we synthesize and categorize this research field in the following dimensions: 1) approaches to augmenting reality; 2) characteristics of robots; 3) purposes and benefits; 4) classification of presented information; 5) design components and strategies for visual augmentation; 6) interaction techniques and modalities; 7) application domains; and 8) evaluation strategies. We formulate key challenges and opportunities to guide and inform future research in AR and robotics.

5 State-of-the-Art Groups

5.1 Interaction & HCI

Caroline Appert (University Paris-Saclay – Orsay, FR), Raimund Dachselt (Technische Universität Dresden, DE), Ricardo Langner (Technische Universität Dresden, DE), and Can Liu (City University – Hong Kong, HK)

This working group conducted a quick survey of the current state of the art for interacting with visualizations. Looking from an HCI perspective on what has been researched in the past, we first identified several interaction modalities, including, but not limited to, mouse and keyboard, touch and multitouch, pen, tangibles, eye gaze, mid-air gestures (hand), spatial interaction and device gestures, embodied interaction, speech, twisting, bending, deformation, proxemic interaction, and any multimodal combination thereof. Many of these modalities have already been applied in visualization research, especially in the past 15 years, with an increasing number of papers devoted to natural interaction or Post-WIMP interaction for visualization, sometimes characterized as visualization beyond the desktop. We provided several examples from the literature for each category, identified gaps, and discussed the potential. Touch input on arbitrarily shaped surfaces is still not easy, pen interaction has rarely been applied to curved surfaces, speech and multimodal interaction including speech would be a good candidate for interacting with unusual displays, and twisting, bending, deformation, etc. are still very underexplored in vis research. This situation will likely change with non-flat, non-rectangular displays being increasingly used for data visualization.

5.2 Perception, Design, and Vis

Lyn Bartram (Simon Fraser University – Surrey, CA), Anastasia Bezerianos (Université Paris-Saclay, INRIA, CNRS, FR), Tanja Blascheck (Universität Stuttgart, DE), Pierre Dragicevic (INRIA – Bordeaux, FR), Petra Isenberg (INRIA Saclay – Orsay, FR), and Bongshin Lee (Yonsei University – Seoul, KR)

This group presented trends in visualization on "atypical" display forms, and atypical use (atypical for visualization research and design) to inform participants from outside the visualization community about relevant work related to the seminar. We noted existing visual perception work on very large displays (walls) and very small ones (smartwatches) that can be applied to study novel form factors for visualization. However, we also highlighted atypical "use",including aspects that go beyond classic visualization tasks, designs, and evaluation methodologies, such as affective responses not only due to the visual encoding but also due to the display form and context of visualization use; aspects related to glanceability (mostly from work on smartwatches), etc. We additionally summarized situations in which visualizations have appeared on atypical "forms." Some are extensions of classic displays, like combining mobile phones and tablets or combining them with digital tables to create new forms and configurations for data visualization, adding flexible displays as bands on smartwatches and extending the visualization surfaces, etc. Others are forms rarely used for visualization, such as drone displays that can fly, spherical displays (from big room installations to small hand-size spheres), flexible or pin-array displays used to visualize data, ambient displays embedded in the environment, and more generally situated visualizations.

5.3 Data Physicalization & Shape-Changing Displays

Lora Oehlberg (University of Calgary, CA), Nathalie Bressa (Télécom Paris, FR), Aluna Everitt (University of Canterbury – Christchurch, NZ), Yvonne Jansen (CNRS – Talence, FR), Konstantin Klamka (Barkhausen Institut – Dresden, DE), Charles Perin (University of Victoria, CA), Kim Sauvé (University of the West of England – Bristol, GB), and Miriam Sturdee (University of St Andrews, GB)

We presented a summary of relevant resources on data physicalizations and shape-changing displays. Data physicalizations (DP) are physical representations of data; shape-changing displays (SCD) are dynamic physical displays that can represent varying types of data or information. We summarized research venues that approach these topics from the perspectives of visualization, human-computer interaction, and enabling technologies (systems engineering). We also highlight communities beyond academia that apply or practice data physicalization or shape-changing displays within their domain. We then discussed relevant research publications that address key research questions, including: what the design space of DP and SCD looks like; how to design and build DP and SCD; what are the "platforms" that designers use to define data physically; how we encode data physically; how people interact with physical properties; what is the role of the audience of a DP or SCD. Finally, we discussed example systems and enabling technologies that make data physicalizations and shape-changing displays possible, including digital fabrication approaches and meta-materials.

5.4 AR/MR/XR Visualization

Tobias Isenberg (INRIA Saclay – Orsay, FR), Emmanuel Pietriga (INRIA Saclay – Orsay, FR), Ryo Suzuki (University of Colorado – Boulder, US & Tohoku University – Sendai, JP), Wesley J. Willett (University of Calgary, CA), and Julie Williamson (University of Glasgow, GB)

This group presented a summary of related work on mixed/augmented/extended reality (XR) visualization approaches and enabling technologies. We first summarized the visualization, immersive technology, and computer graphics venues around which this research is centered. Next we outlined a slate of recent innovations in XR hardware that have substantially reduced barriers to creating rich, complex and usable XR visualization systems, including: (1) untethered headsets, (2) table tracking in real-world settings, (3) high-resolution video passthrough, (4) fine-grained hand tracking, and (5) live environment mapping. We then highlighted a variety of recent design spaces and example systems that represent the state-of-the-art in XR visualization.

6 Grand Challenges Working Groups

Identifying grand-challenges in emerging research topics, such as visualization on non-flat displays, is crucial to help identify and articulate the most pressing questions and guide future research efforts. To inspire discussions and reflection, we first conducted a series of design futuring exercises, using an established ideation methodology (Vis Future Cards). These produced a suite of novel designs for visualizations built for shape-changing, evolving, wearable, and natural material displays; as well as discussions around opportunities such displays provide for data communication, sharing, and engaging with data. Inspired by these reflections and the prior work, the seminar participants then conducted a series of brainstorming sessions about challenges, opportunities and open research questions. We then thematically analyzed the resulting ideas and grouped them in categories that represent the major challenges in the domain. Breakout working groups explored these topics and started creating an analysis that we plan to publish as a grand challenges / research agenda report. In the following sections, we provide abstracts for these working groups.

These include: determining the scope of research on the domain; exploring display technologies and interaction techniques that may fit the display form forms; identifying what visualizations and tasks such displays will be best used for (application contexts); considering the tight relationship between data and display and how they can each influence the other; adapting existing visualization design and prototyping methodologies for displays that are non-flat and whose form may need to fit the environment; and finally evaluation challenges and opportunities to determine how we read but also how we engage with visualizations on them given that these displays do not only encode data but also need to fit their environment.

6.1 What is the scope of visualization on non-flat non-rectangular displays (NFNRD)?

Julie Williamson (University of Glasgow, GB), Nathalie Bressa (Télécom Paris, FR), Pierre Dragicevic (INRIA – Bordeaux, FR), Yvonne Jansen (CNRS – Talence, FR), Konstantin Klamka (Barkhausen Institut – Dresden, DE), Lora Oehlberg (University of Calgary, CA), Emmanuel Pietriga (INRIA Saclay – Orsay, FR), Kim Sauvé (University of the West of England – Bristol, GB), Miriam Sturdee (University of St Andrews, GB), Ryo Suzuki (University of Colorado – Boulder, US & Tohoku University – Sendai, JP), and Wesley J. Willett (University of Calgary, CA)

Recent years have seen an increase of visualization research that goes beyond desktop displays, this include visualization on mobile devices, physicalization, and immersive visualization using augmented reality technology. We focused on articulating the difference of the work targeted in this seminar from that prior work. Our focus targets dynamic and interactive visualizations, on physical displays that are not flat but rather have forms that integrate / fit their environment. These characteristics suggest that such displays allow embedding data in context (for example, in locations where the data is relevant) and can support more expressive, immediate, and powerful interactions. And they allow experiencing data in person rather than via a window into another place, as is the case of flat displays. Other work groups elaborated on the benefits of such displays as well as open challenges and opportunities. After defining the nature of these displays and visualizations built for them, we then considered the challenges they pose. These include how to promote displays and visualization use cases that minimize material and other resources (specialized material which may limit their life-cycle, energy use, sustainability, development cost), as well as ethical considerations (for example, ensuring equity, access, privacy, etc.).

6.2 Display and input technologies

Caroline Appert (University Paris-Saclay – Orsay, FR), Pierre Dragicevic (INRIA – Bordeaux, FR), Tobias Isenberg (INRIA Saclay – Orsay, FR), Ricardo Langner (Technische Universität Dresden, DE), Can Liu (City University – Hong Kong, HK), and Ryo Suzuki (University of Colorado – Boulder, US & Tohoku University – Sendai, JP)

We identified possible broad categories of displays that fit our scope of non-flat non-rectangular displays. These include real 3D displays (such as volumetric systems or 3D LED arrays), non-deformable [non-flat or non-rectangular] displays, deformable displays (including actuated, bendable, and foldable variants), modular (compound) displays made up of several components that can be rearranged, wearable or fabric displays, as well as everyday objects as displays. In addition to these form factors, we consider challenges related to materials, resolution, durability and energy consumption. The interactions appropriate for these displays are very much an open research topic. We have identified several input technologies that either exist or can be envisioned for such displays, such as on-surface input (via touch or tools) or off-surface input (mid-air gestures, speech, or proxemics). A first open research question is categorizing these possible inputs and studying good matches between input and type of display. Nevertheless, the details of how input is going to be translated to "interactions" (the words or commands that target and manipulate visualizations) depends on the type of visualizations and data presented on them.

6.3 Applications: What is the benefit of visualizing data on NFNRD?

Tanja Blascheck (Universität Stuttgart, DE), Caroline Appert (University Paris-Saclay – Orsay, FR), Aluna Everitt (University of Canterbury – Christchurch, NZ), Petra Isenberg (INRIA Saclay – Orsay, FR), Tobias Isenberg (INRIA Saclay – Orsay, FR), and Can Liu (City University – Hong Kong, HK)

As NFNRDs can vary greatly in form and size, identifying when data visualization on NFNRDs makes sense (and what benefits it might provide) requires taking a step back and considering the possible contexts of use. We considered several dimensions that define this context of use: application types (public / private, stationary /mobile), audience (data specialists / general public, single person / collaborative use), tasks (ambient / focus), etc. Clearly not all display types (identified in the previous working group) fit all contexts of use. We considered possible matches and identified a grand challenge focued on the interplay between application (purpose and context of the visualization), data, and displays.

6.4 Relationship between data and displays

Anastasia Bezerianos (Université Paris-Saclay, INRIA, CNRS, FR), Lyn Bartram (Simon Fraser University – Surrey, CA), Raimund Dachselt (Technische Universität Dresden, DE), Konstantin Klamka (Barkhausen Institut – Dresden, DE), Ricardo Langner (Technische Universität Dresden, DE), Charles Perin (University of Victoria, CA), and Julie Williamson (University of Glasgow, GB)

Having access to displays that can be of diverse forms and sizes raises several research questions around how to display data on them. These start from visualization mapping and resolution questions – what we call "how the data follows the display". For example, how can visualizations designed for a flat rectangular screen be rendered on other types of displays, such as a sphere or a cylinder? More generally, can we render any visualization on any display form and, if not, what are the limits we encounter (resolution, occlusion, etc)? Other questions relate to how the nature of the data can influence the display, or "how the display follows the data". These observations highlight how designers might allow data to shape the form of actuated or shape-changing displays. We articulated the connection between data and displays and collected a set of opportunities and research challenges on this topic.

6.5 Design approaches for NFNRD visualizations

Kim Sauvé (University of the West of England – Bristol, GB), Nathalie Bressa (Télécom Paris, FR), Raimund Dachselt (Technische Universität Dresden, DE), Lora Oehlberg (University of Calgary, CA), Charles Perin (University of Victoria, CA), and Miriam Sturdee (University of St Andrews, GB)

But how do we go about designing visualizations for such displays? Traditional sketching and paper prototyping methods are not enough, as they do not capture the subtleties of displaying information on forms that may distort visuals (see also subsection 6.6) or may be hard to envision on a two dimensional medium such as paper. We collected factors that make NFNRD hard to design for and considered different existing design methods, such as physical prototyping, and how these could be adapted for creating initial designs. We additionally reflected on the use of these displays, as their purpose is to be more integrated than flat displays, exploring how they could be perceived in their intended context of use is important. And thus how our design methods could ensure the context of use is taken into account even at the early stages.

6.6 Evaluating visualizations on NFNRD

Lyn Bartram (Simon Fraser University – Surrey, CA), Anastasia Bezerianos (Université Paris-Saclay, INRIA, CNRS, FR), Tanja Blascheck (Universität Stuttgart, DE), Petra Isenberg (INRIA Saclay – Orsay, FR), and Emmanuel Pietriga (INRIA Saclay – Orsay, FR)

Given the diverse and unpredictable form factors and sizes of NFNRD, it is important to evaluate how visualizations are read on them. The topic of evaluation is challenging as there are several factors that need to be considered – including the form of the display and the context of use. The form needs to be considered from a readability / perception perspective, in other words we need to validate if information can be correctly read and understood, as aspects like curvature, angle, etc. may affect visualization reading. Traditional user-study methods will likely be enough in these instances. But we additionally need to consider evaluating the context of use and of display placement. It is possible that data engagement will be higher with displays of unusual form factors or form factors that have forms fitting the environment. Investigating and evaluating engagement and affective responses requires different study methods, such as longitudinal observation studies or controlled studies that capture measures beyond time and error. Our group listed a set of relevant study methods, measures, and logistics

7 Outlook and Conclusion

Throughout the week, our group grappled with the vast scale, potential, and implications for visualization as new display technologies allow us to escape the narrow confines and constraints of 2D screens. The diversity and heterogeneity of future display technologies that might support data visualization is vast and complex – to the extent that they simple generalizations. Yet within this rich space, our seminar succeeded in identifying not only a range of clear challenges and opportunities for the visualization research community, but also a set of promising near-term research projects (which smaller teams of attendees intend to pursue going forward). We look forward to a set of inspired new publications and research grants which trace their genesis to this seminar, and extend a gracious thank you to Dagstuhl for providing the support and organization that allowed us to convene it.

Participants

- Caroline Appert
University Paris-Saclay – Orsay, FR
- Lyn Bartram
Simon Fraser University – Surrey, CA
- Anastasia Bezerianos
Université Paris-Saclay, INRIA, CNRS, FR
- Tanja Blascheck
Universität Stuttgart, DE
- Nathalie Bressa
Télécom Paris, FR
- Raimund Dachselt
TU Dresden, DE
- Pierre Dragicevic
INRIA – Bordeaux, FR
- Aluna Everitt
University of Canterbury – Christchurch, NZ
- Petra Isenberg
INRIA Saclay – Orsay, FR
- Tobias Isenberg
INRIA Saclay – Orsay, FR
- Yvonne Jansen
CNRS – Talence, FR
- Konstantin Klamka
Barkhausen Institut – Dresden, DE
- Ricardo Langner
TU Dresden, DE
- Bongshin Lee
Yonsei University – Seoul, KR
- Can Liu
City University – Hong Kong, HK
- Lora Oehlberg
University of Calgary, CA
- Charles Perin
University of Victoria, CA
- Emmanuel Pietriga
INRIA Saclay – Orsay, FR
- Kim Sauvé
University of the West of England – Bristol, GB
- Miriam Sturdee
University of St Andrews, GB
- Ryo Suzuki
University of Colorado – Boulder, US & Tohoku University – Sendai, JP
- Wesley J. Willett
University of Calgary, CA
- Julie Williamson
University of Glasgow, GB

Report from Dagstuhl Seminar 25091

Tradeoffs in Reactive Systems Design

Jerónimo Castrillón-Mazo*[1], Chadlia Jerad*[2], Edward A. Lee*[3], Claire Pagetti*[4], and Shaokai Jerry Lin†[5]

1 **TU Dresden, DE.** jeronimo.castrillon@tu-dresden.de
2 **University of Manouba, TN.** chadlia.jerad@ensi-uma.tn
3 **University of California – Berkeley, US.** eal@berkeley.edu
4 **ONERA – Toulouse, FR.** claire.pagetti@onera.fr
5 **University of California – Berkeley, US.** shaokai@berkeley.edu

Abstract

Reactive systems – software systems that continuously interact with their physical or digital environment – are central to safety-critical domains such as autonomous vehicles, industrial automation, and medical devices. These systems face inherent design tensions: the need to be predictable yet adaptable, timely yet accurate, consistent yet available, and secure yet accessible. Addressing one requirement often undermines another, revealing tradeoffs that are not merely engineering challenges but fundamental limits. This seminar aimed to confront these tradeoffs directly, drawing on insights from real-time systems, distributed computing, formal methods, machine learning, and security. By exploring case studies, formal frameworks, and practical tools, we made progress in the understanding of how to make design decisions when no single solution satisfies all competing goals. Interactive sessions and demos gave participants a tangible sense of the costs and compromises involved. With all this, we achieved the seminar's goal, namely, to cultivate a shared understanding of how to navigate the complex design space of reactive systems and chart paths toward more robust and principled solutions.

Seminar February 23–28, 2025 – https://www.dagstuhl.de/25091

2012 ACM Subject Classification Computer systems organization → Embedded and cyber-physical systems; Computer systems organization → Dependable and fault-tolerant systems and networks; Computer systems organization → Real-time systems; Computing methodologies → Distributed computing methodologies; Computing methodologies → Concurrent computing methodologies; Mathematics of computing → Mathematical analysis; Security and privacy → Software and application security

Keywords and phrases reactive systems, cyber-physical systems, design tradeoffs, real-time systems, distributed computing, predictability, adaptability, timeliness, accuracy, consistency, availability, security, accessibility, machine learning, formal methods, system design, embedded systems, safety-critical systems, tool support, programming models, runtime verification

Digital Object Identifier 10.4230/DagRep.15.2.126

* Editor / Organizer
† Editorial Assistant / Collector

1 Executive Summary

Jerónimo Castrillón-Mazo (TU Dresden, DE)
Chadlia Jerad (University of Manouba, TN)
Edward A. Lee (University of California – Berkeley, US)
Claire Pagetti (ONERA – Toulouse, FR)

Reactive systems are software systems that engage in a continual dialogue with their environment. They constitute the software parts of cyber-physical systems where timely reactions are often critical to safety. Applications include autonomous vehicles, electric power systems, industrial automation, healthcare electronics, and robotics. Because the software engages in a continual dialog with its environment, it often has conflicting requirements. It needs to be predictable, but robust to unpredictable events; it needs to react in a timely manner, but this often requires reacting with inconsistent information; it needs to be adaptable, but demonstrably safe; and it needs to be secure, but accessible and available. Many conferences and workshops focus on one of the goals, such as achieving real-time behavior, without explicitly acknowledging the costs and without providing sound strategies for dealing with failures that prevent reaching the goals. The focus of this seminar was on the tradeoffs that are intrinsic in the design of such systems. When you make a system predictable, available, secure, or even demonstrably safe, what have you lost?

This seminar pulled in experts from manifold disciplines, both academic and industrial, to identify and discuss the fundamental limits in reactive systems design that make tradeoffs inevitable. In preparation for the seminar, the co-organizers reached out to leading experts among the participants and invited them to deliver four talks to frame the discussions for each of the first four days. Following this initial outreach, all participants were contacted and invited to contribute through short talks, position statements, and demonstrations of any relevant tools. The seminar format was kept flexible and open to allow space for ideas and key discussion points to emerge organically. In this context, the working groups described in Section 4 arose from the core ideas and challenges identified during the morning sessions. These groups held their discussions during the first half of the afternoon sessions of the first two days. The seminar was organized as described blow.

Day 1: Consistency vs. availability

Consistency is agreement on shared information across a distributed system. Availability is the ability to act on that shared information in a timely way. It has been shown that as latency increases when sharing such information, either consistency or availability or both must be sacrificed. This topic focused on how to manage this tradeoff. It included presentations proposing different ways to formalize the tradeoff, as well as examples of how it arises in software design. Two breakout groups formed on the first day, focusing on the topics “Distributed Music Challenge” and “Can AI Be Used in Critical Systems?”

Day 2: Timeliness vs. accuracy

Because reactive systems interact continuously with their environment, they need to sense and interpret that environment. Today, many such systems need to include sophisticated vision subsystems, audio information processing, motion sensing, etc. The computation required to interpret sensor data often implies unacceptable delays or impossible energy requirements. It is not acceptable for an automated vehicle to identify a pedestrian after it has hit the pedestrian. This topic focused on how to manage this tradeoff. Three breakout groups formed on the second day, focusing on the topics "Benchmarks for RT systems", "Tradeoff of timeliness and accuracy", and "Orchestration/coordination languages vs reactive languages."

Day 3: Predictability vs. adaptability

Reactive systems often perform critical tasks. We need for them to behave predictably during normal operation, but also adapt to behave reasonably in abnormal situations. Recent innovations in machine learning promise significant improvements for the latter requirement, but it is unclear how to reconcile the use of ML with the former requirement. This topic focused on how to manage this tradeoff.

Day 4: Security vs. accessibility

When systems are secure, nothing bad happens even when malicious players are present. Achieving the goal that "nothing bad happens," however, is trivially easy by ensuring that nothing at all happens. Security measures often get in the way of other goals. This topic will focus on how to manage this tradeoff. For example, techniques that offer tiered access to capabilities, taint analysis, or mixtures of encrypted and unencrypted communication might be explored. During the afternoon session, reports from the group discussions were shared, along with brief previews of the software demonstrations (teasers) scheduled for Day 5.

Day 5: Tools and Demos

This topic focused on tools that support analysis and design and make explicit the management of tradeoffs. The key goal of the groups was to feel the pain intrinsic to the tradeoffs that are the theme of the seminar. A total of five software tools tutorials were presented (see Section 5). Hands-on exercises were organized into two sessions, each consisting of parallel tracks. This structure was designed to give participants the opportunity to experiment with two of the five available tools, rather than one. The tools presented were: "Lingua Franca", "Timed SCCharts", "QRML", "Rebecca", and "HipHop".

2 Table of Contents

Working groups

3 Overview of Talks

3.1 Safety and AI Standards for Automobile: An overview

Andres Barrilado (NXP Semiconductors – Toulouse, FR)

Joint work of Andres Barrilado, Iban Guinebert

Presentation in which a short summary of recently published safety standards for AI in the automotive domain is shared. Some selected extracts are discussed with regards to tradeoffs linked to predictability and adaptability. Additionally, we provide an introduction to the tradeoffs now being observed specifically by semiconductor manufacturers when having to produce functionally safe inference accelerators.

3.2 Modeling reconfigurable CPS using UML

Grzegorz Bazydło (University of Zielona Gora, PL)

The presentation introduces an approach to designing reconfigurable cyber-physical systems (CPS) using state machine diagrams from the Unified Modeling Language (UML). The UML specification is transformed using model-driven development (MDD) techniques into an efficient hardware description language (HDL) program, utilizing a concurrent finite state machine (CFSM) as an intermediate model. The resulting HDL specification can be analyzed, validated, synthesized, and ultimately implemented in field-programmable gate array (FPGA) devices.

Dynamic partial reconfiguration – a feature of modern FPGAs – enables the replacement of parts of the CPS algorithm without powering down the device. However, to leverage this feature, the model must be safe, which, in the proposed approach, means incorporating special idle states where control is transferred during the reconfiguration process. The CFSM model significantly facilitates this task.

The proposed design method provides an efficient graphical modeling approach for the control part of a reconfigurable CPS, along with an automated translation of the behavior model into a synthesizable Verilog description. This Verilog code can be directly implemented in FPGA devices and dynamically reconfigured as needed. A practical example from the field of demand-side management illustrates the successive stages of the proposed method.

3.3 Reactive systems: Optimizations and opportunities in domain-specific computing

Jerónimo Castrillón-Mazo (TU Dresden, DE)

This talk discusses optimizations for dataflow and reactive programs. This includes optimizations to the runtime system of Lingua Franca [5], initial work on decoupling timelines to expose more parallelism [6], a system to explore energy-performance tradeoffs [7, 2], and a methodology to explore tradeoffs between performance and accuracy for ML workloads [3, 4]. The talk closes with an outlook on how domain-specific programming abstractions and domain-specific computer architectures [1] may offer more control over system-level tradeoffs, including performance, energy consumption, accuracy and reliability.

References

1 Asif Ali Khan, Hamid Farzaneh, Karl F. A. Friebel, Lorenzo Chelini, and Jeronimo Castrillon. Cinm (cinnamon): A compilation infrastructure for heterogeneous compute in-memory and compute near-memory paradigms. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'25)*, ASPLOS '25. Association for Computing Machinery, March 2025.
2 Robert Khasanov and Jeronimo Castrillon. Energy-efficient runtime resource management for adaptable multi-application mapping. In *Proceedings of the 2020 Design, Automation and Test in Europe Conference (DATE)*, DATE '20, pages 909–914. IEEE, March 2020.
3 Guilherme Korol, Michael Guilherme Jordan, Mateus Beck Rutzig, Jeronimo Castrillon, and Antonio Carlos Schneider Beck. Design space exploration for CNN offloading to FPGAs at the edge. In *2023 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, pages 1–6, Los Alamitos, CA, USA, June 2023. IEEE, IEEE Computer Society.
4 Guilherme Korol, Michael Guilherme Jordan, Mateus Beck Rutzig, Jeronimo Castrillon, and Antonio Carlos Schneider Beck. Pruning and early-exit co-optimization for CNN acceleration on FPGAs. In *Proceedings of the 2023 Design, Automation and Test in Europe Conference (DATE)*, DATE'23, pages 1–6. IEEE, April 2023.
5 Christian Menard, Marten Lohstroh, Soroush Bateni, Mathhew Chorlian, Arthur Deng, Peter Donovan, Clément Fournier, Shaokai Lin, Felix Suchert, Tassilo Tanneberger, Hokeun Kim, Jeronimo Castrillon, and Edward A. Lee. High-performance deterministic concurrency using Lingua Franca. *ACM Transactions on Architecture and Code Optimization (TACO)*, 20(4):1–29, August 2023.
6 Julian Robledo, Christian Menard, Erling Jellum, Edward A. Lee, and Jeronimo Castrillon. Timing enclaves for performance in Lingua Franca. In *2024 Forum for Specification and Design Languages (FDL)*, pages 1–9, September 2024.
7 Till Smejkal, Robert Khasanov, Jeronimo Castrillon, and Hermann Härtig. E-Mapper: Energy-efficient resource allocation for traditional operating systems on heterogeneous processors, 2024.

3.4 Formalizing Tradeoffs in Reactive Systems Design

Samarjit Chakraborty (The University of North Carolina (UNC) at Chapel Hill, US)

Today, most complex reactive systems, and especially autonomous systems, are composed of multiple subsystems such as controllers, schedulers, machine learning (ML) components, and security systems. These subsystems are designed independently and are aimed to work to perfection. But in reality they do not work perfectly and the imperfection is assumed to be small enough that it may be ignored. We argue that design approaches should make the imperfections in these components first class citizens, formalize specifying imperfections, and allow tradeoffs in the design play a more prominent role.

Towards this, we first define a flexible notion of system-level safety to admit imperfect behaviors. We then compute which component behaviors would ensure such a notion system safety. This allows more component behaviors compared to those that would be admitted if the component was to be designed to perfection. These additional behaviors allow more efficient component implementations. Further, they allow exploration of tradeoffs where the deficiency of one component may be compensated by another component, while satisfying the notion of system-level safety and the flexibility it allows. To arrive at a general notion of imperfection, we consider the dynamics of closed-loop control systems. First, we determine the trajectory followed by such systems in their state space, when all the system components behave perfectly, *e.g.*, when all the elements of the system state may be estimated exactly, all machine learning components in the system always return accurate inferences, and the deadlines of all software/control tasks are always met. For any given initial condition, such a system trajectory represents the *ideal* behavior of the system, which might not be practically realizable. To allow imperfect system behaviors, we consider a *safety pipe* around this ideal trajectory. Any system trajectory that lies within this safety pipe is considered to be an acceptable or safe behavior of the system [4].

It is, however, non-trivial to estimate component behaviors from such specifications of safety pipes that capture system-level behaviors. In this talk we illustrated how this may be done for the specific case of imperfect timing behaviors. In particular, we asked: Given a safety pipe specifying allowed system behaviors, what are feasible timing behaviors that a control task may experience? Our procedure to answer this question involves the following two steps: (1) Guess a regular language $\mathcal{L}$ of timing behaviors that a control task may experience [1]. Such a language is a set of strings over the alphabet $\Sigma = \{0, 1\}$. Here, 0 represents the case where the control task in question misses its deadline at the specified sampling period, *i.e.*, the control input is not available at the time of actuation and an old control input is used. On the other hand, 1 denotes the case that the control task successfully completes its execution before the predetermined time of actuation. Hence, a string $101010\ldots$ denotes that the case where every alternate execution of the control task misses its deadline. Equivalently, it also denotes the case where the control task is purposefully not scheduled in every alternate sampling period to save computation bandwidth.

In the next step: (2) We use approximate reachability analysis over a bounded time horizon to check whether all the trajectories of the given closed loop system stay within the specified safety pipe, when the control task is subjected to any timing behavior that belongs to the language $\mathcal{L}$. Here, we use *approximate* reachability to contain the state-space explosion that would otherwise happen. If the language $\mathcal{L}$ of timing behaviors result in safe closed-loop behaviors (*i.e.,* all the closed-loop trajectories are within the safety pipe) then we add $\mathcal{L}$ to the set of safe languages, *viz.*, $\mathcal{L}_{safe} = \mathcal{L}_{safe} \cup \mathcal{L}$, where $\mathcal{L}_{safe}$ was initially empty. By iteratively following this procedure, we create a set of timing behaviors $\mathcal{L}_{safe}$ that results in safe closed-loop system behaviors. Now, since the union of regular languages is regular, $\mathcal{L}_{safe}$ may be represented as a finite state automata. We can use this procedure to synthesize safe schedules for a set of control tasks. For this, we rely on automata-theoretic techniques [5, 9] to combine the $\mathcal{L}_{safe}$s of all the controllers to synthesize schedules where the tasks of each controller is not always scheduled for execution (*i.e.*, occasionally misses its deadline), but the safety of all the closed-loop systems is guaranteed [10]. Such schedules account for the dynamics of the closed-loop system [2, 3, 6] and cannot be reproduced by standard scheduling policies like fixed priority or earliest deadline first (EDF) [7]. This approach of using a flexible notion of system safety to allow multiple (imperfect) component behaviors allows the exploration of different tradeoffs between the performance of different system components. For example, the work in [11] uses the size of the reachable set as a measure of safety and shows how the size of a neural network – and therefore its inference quality – for state estimation in a closed-loop control system may be traded off with the magnitude of the control input. This is used to partition a deep neural network (DNN) between edge and cloud resources. A similar approach was also used [8] to purposefully choose DNNs with less than optimal size and inference accuracy, to fit multiple DNNs on the same shared graphics processing unit (GPU), while ensuring that the dynamics of the closed-loop systems using these DNN inferences remain safe.

While a safety pipe around an ideal system trajectory, or the size of the reachable set, are two different notions of system safety, there are many other possibilities. For example, not all trajectories within a safety pipe might be permissible. Most work on formal methods till date have focused on specifying perfect or ideal system behaviors (*e.g.*, all deadlines are met). Specifying what kind of imperfect behaviors might be acceptable requires domain knowledge, and those with such knowledge are usually not sufficiently conversant with formal methods to be able to write formal specifications. Hence, new avenues for involving domain experts – who are not specialists in formal methods – to write formal specifications, perhaps using tools from machine learning will be needed.

References

1 Bineet Ghosh, Clara Hobbs, Shengjie Xu, F. Donelson Smith, James H. Anderson, P. S. Thiagarajan, Benjamin Berg, Parasara Sridhar Duggirala, and Samarjit Chakraborty. Statistical verification of autonomous system controllers under timing uncertainties. *Real Time Syst.*, 60(1):108–149, 2024.

2 Dip Goswami, Reinhard Schneider, and Samarjit Chakraborty. Co-design of cyber-physical systems via controllers with flexible delay constraints. In *16th Asia South Pacific Design Automation Conference (ASP-DAC)*, 2011.

3 Dip Goswami, Reinhard Schneider, and Samarjit Chakraborty. Re-engineering cyber-physical control applications for hybrid communication protocols. In *Design, Automation and Test in Europe (DATE)*, 2011.

4 Clara Hobbs, Bineet Ghosh, Shengjie Xu, Parasara Sridhar Duggirala, and Samarjit Chakraborty. Safety analysis of embedded controllers under implementation platform timing uncertainties. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, 41(11):4016–4027, 2022.

5 Clara Hobbs, Shengjie Xu, Bineet Ghosh, Enrico Fraccaroli, Parasara Sridhar Duggirala, and Samarjit Chakraborty. Quantitative safety-driven co-synthesis of cyber-physical system implementations. In *15th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*, pages 99–110. IEEE, 2024.

6 Martin Lukasiewycz, Reinhard Schneider, Dip Goswami, and Samarjit Chakraborty. Modular scheduling of distributed heterogeneous time-triggered automotive systems. In *17th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2012.

7 Florian Sagstetter, Sidharta Andalam, Peter Waszecki, Martin Lukasiewycz, Hauke Stähle, Samarjit Chakraborty, and Alois C. Knoll. Schedule integration framework for time-triggered automotive architectures. In *The 51st Annual Design Automation Conference (DAC)*, 2014.

8 Shengjie Xu et al. GPU partitioning & neural architecture sizing for safety-driven sensing in autonomous systems. In *IEEE International Conference on Assured Autonomy (ICAA)*, 2024.

9 Shengjie Xu, Bineet Ghosh, Clara Hobbs, P. S. Thiagarajan, and Samarjit Chakraborty. Safety-aware flexible schedule synthesis for cyber-physical systems using weakly-hard constraints. In Atsushi Takahashi, editor, *28th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2023.

10 Anand Yeolekar, Ravindra Metta, Clara Hobbs, and Samarjit Chakraborty. Checking scheduling-induced violations of control safety properties. In *20th International Symposium on Automated Technology for Verification and Analysis (ATVA)*, volume 13505 of *Lecture Notes in Computer Science*. Springer, 2022.

11 Tingan Zhu, Prateek Ganguli, Arkaprava Gupta, Shengjie Xu, Luigi Capogrosso, Enrico Fraccaroli, Marco Cristani, Franco Fummi, and Samarjit Chakraborty. Controllers for edge-cloud cyber-physical systems. In *17th IEEE International Conference on COMmunication Systems and NETworks (COMSNETS)*, 2025.

3.5 Security Efficiency Tradeoffs for Intelligent Systems

Anupam Chattopadhyay (Nanyang TU – Singapore, SG)

Joint work of Anupam Chattopadhyay, Prasanna Ravi, Sourav Sen Gupta, Mustafa Khairallah, Zakaria Najm, Shivam Bhasin, Arpan Jati, Naina Gupta, Anh Tu Ngo

Security and privacy are growing concerns across all forms of digital systems, including Cyber-Physical Systems (CPS). Embedding intelligence in such systems are done to enhance performance and autonomy. Such an inclusion compromises security, which is often fixed as an afterthought. The talk highlights several instances across the stack of digital systems, from machine learning, general purpose processors to hardened cryptographic implementations, which showcase that security and efficiency present contrasting choices for a system designer. This situation is exacerbated for autonomy, where Artificial Intelligence (AI) is invoked. Lack of explanation in many AI systems directly falls into the hands of a malicious actor, who can compromise the system in many subtle ways.

3.6 Assurance of Neural Network-based Safety-Critical Avionics with Formal Methods

Arthur Clavière (Collins Aerospace – Blagnac, FR)

Joint work of Arthur Clavière, Dmitrii Kirov, Darren Cofer

We present a process for the Verification of neural networks using formal methods, developed to support the needs of Collins engineers working on neural networks for aircraft. The process focuses on low-complexity neural networks (having hundreds or thousands of learnable parameters) with simple input domains (less than 10 input features). It is especially applicable when the neural network is used to approximate a given target function (such as a complex optimization or large look-up table) at a lower computational cost. It compares the behavior of the neural network to the target function by exhaustively verifying the entire input space. Such an exhaustive verification, made possible by using formal methods, analyzes how the neural network would respond not only on the training or test dataset but also to any unseen data in its input range. We anticipate that this approach will be useful for high-criticality neural network-based components (DAL-C and above). It can be used as a means of compliance with certification objectives e.g., stability analysis and generalization capability assessment as defined by the European Union Aviation Safety Agency (EASA). The process is illustrated through a Collins use case, namely the Recommended Cruise Level (RCL) function, which can be used to provide pilots with a suggested altitude to reduce the operational cost of the flight. The process is used to verify a neural network implementation of the RCL, that significantly reduces the computational cost and memory footprint of a traditional implementation.

3.7 A component model and modelling language for tradeoffs and multi objective optimization

Marc Geilen (TU Eindhoven, NL)

Joint work of Marc Geilen, Twan Basten

Tradeoffs represent choices between multiple objectives that are determined by choices between different configurations. This presentation introduces a mathematical framework, a component model and a domain-specific language. The mathematical framework captures the essential operations and requirements to support compositional computation of Pareto optimal configurations of a system from descriptions of the (optimal) configurations of components. It defines a number of relations on configurations, the sets of possible configurations of components or systems to characterize which ones are considered better, worse, or equivalent. The framework provides insight in requirements on the models and operations to work effectively, preserve optimality, etcetera. We further introduce a component model for Quality and Resource Management that builds on Pareto Algebra. It aims to describe application and platform components in a virtualized platform setting. Components are described in terms of input and out qualities, their required and provided budgets, their optimization objectives and the parameters used to configure the component. The resulting system semantics is a set of (Pareto optimal) configurations of the overall system. This set can be implicitly captured as a set of equations in terms of the combined system parameters. We have defined a Domain-Specific Language, called QRML, to specify such component models generated different types of visualization and generate the set of constraints to be checked for satisfiability, or for optimization, by a constraint solver, such as the Z3 SMT solver. The tools are available through the web page `https://www.qrml.org`.

3.8 Exploring the memory / execution time tradeoff in dataflow graphs

Alain Girault (INRIA – Grenoble, FR)

Joint work of Alain Girault, Pascal Fradet, Alexandre Honorat

Many computing systems are constrained by a fixed amount of available shared memory. Modeling applications with task graphs makes it possible to analyze and optimize their memory usage. The NP-complete problem studied here is finding a parallel schedule of a given task graph that minimizes its memory peak.

Our first contribution is an algorithm that finds optimal sequential schedules for a dataflow task graph. This algorithm is based on graph transformation rules. On a large class of graphs, it is able to compress the graph into a single node which contains a sequential schedule optimal w.r.t. the memory peak. The approach also applies to SDF graphs after converting them into task graphs. However, since that conversion may produce very large graphs, we also propose a new suboptimal method, similar to Partial Expansion Graphs, to reduce the problem size. We evaluate our approach on classic benchmarks, on which we always outperform the state-of-the-art.

From this optimal sequential schedule, our second contribution is a dynamic parallel schedule, which consists of a ready list scheduling that we adapt to take into account memory requirements. Our approach always produces a parallel schedule that meets the constraints and enjoys very good speedups. It can also be applied to less harsh memory constraints, leading to more substantial speedups. We compare with alternative approaches on multiple applications expressed as task graphs (scientific workflows, signal processing filters). When memory constraints are close to their minimum, our approach always succeeds in finding a parallel schedule meeting the given constraints, whereas the other approaches mostly fail. When memory constraints are significantly higher, our results are comparable to others for speedup. Furthermore, our approach is faster and can deal with very large task graphs (up to 50,000 nodes) using a naive Python implementation.

3.9 Latency and Consistency Tradeoffs in Shared-Memory Systems

Andrés Goens Jokisch (University of Amsterdam, NL)

Shared-memory systems are a reality of today's hardware, and current trends are to increase their scale and complexity, with heterogeneous shared-memory architectures and hierarchical systems with protocols like CXL. These systems, however, have complex behaviors that trade consistency with latency, through out-of-order execution and weak guarantees in the protocols, resulting in weak memory consistency models, i.e. models weaker than sequential consistency. In this talk, I present some of the architectural guarantees, and open questions in this context, which give us primitives to express and control tradeoffs in consistency and latency. I argue we should build higher-level abstractions on top of these primitives to reason about these tradeoffs at a higher level of abstraction in the design process, e.g. programming languages or runtime systems.

3.10 Can we make COTS CPS Ultra-Reliable?

Arpan Gujarati (University of British Columbia – Vancouver, CA)

In the avionics domain, "ultra-reliability" refers to the practice of ensuring quantifiably negligible residual failure rates in the presence of transient and permanent hardware faults. In this talk, I discuss the need for new mechanisms that can help us make contemporary and next-generation CPS, which use inexpensive, relatively unreliable off-the-shelf components, more reliable and, if possible, ultra-reliable (as airplanes). I will frame the discussion around the problem of Byzantine fault tolerance (BFT). Specifically, I will briefly present our RTSS 2022 work, where we proposed a novel BFT timed key-value store, called Achal, designed specifically for Ethernet-based distributed real-time control applications. I will then describe our current work on understanding two key limitations of Achal-like systems. First, can they be seamlessly ported to real, complex CPS applications? Second, can precision time synchronization be considered a reliable primitive? Finally, I wrap up with my key takeaway, that we must focus on systems / middleware / tools that make it easier to engineer reliable systems, as opposed to designing new fault-tolerance protocols from scratch.

3.11 Necessary Conditions for Model Engineering to Ensure System Correctness

Jérôme Hugues (Carnegie Mellon University – Pittsburgh, US)

Engineering Reactive Systems primarily involves evaluating the accuracy of computations, as well as the system's quality attributes (such as safety, security), or its functions or data (such as availability, consistency). This list is not exhaustive, and neither is this mapping. These properties can be assessed by testing the final system. A more timely and cost-effective approach is to rely on models and verification and validation (V&V) techniques such as simulation or model checking. Evaluating these quality attributes necessitates the construction of an additional system: a model and a demonstration that the model is sufficiently accurate to both represent the system and support a verification objective.

In this presentation, I propose some considerations in the scope of Dagstuhl Seminar 25091. Firstly, when considering a model as a system, it is crucial to define its requirements. The term "capturing a system behavior" is too ambiguous. A model should support an engineering goal, such as evaluating a system attribute. Therefore, the selection of a modeling language, the creation of the model, and the property to be evaluated cannot be separated. It is essential to ensure that the evaluation of an attribute or a tradeoff analysis among multiple attributes can be initially expressed in a model and then analyzed by some techniques. This can only be achieved if the modeling language is sound, the model captures a domain taxonomy, and, of course, it is correct.

3.12 Introduction to Avionic Certification

Victor Jegu (Airbus S.A.S. – Toulouse, FR)

Before aircrafts are approved to carry passengers, the industry has to provide evidence to the aviation authorities of the safety and conformity of their products. This approval is called "certification". And is achieved by conducting a very thoroughly documented process of addressing every foreseeable failure with safety impact. This process is particularly stringent with failures of catastrophic or hazardous consequences. This talk briefly describes this process for avionic equipment (computers). For functions with minor safety impact, COTS or COTS-like "Design Assurance Level" (DAL) could be considered. At such a level, low energy (edge) AI technology could already be considered. But for higher Assurance Level, achieving this objective may require a thorough understanding of the technology, exhaustive (including formal) design checking, use of diverse strategies for error detection, and resource redundancy. The purpose of this talk is to give academics with intent to assist industry by providing tools, methods, languages and IP in general, an understanding of the level and nature of information (documentation and support) the industry needs to adopt their proposals.

3.13 A Motivating Example: Distributed Music

Erling Rennemo Jellum (University of California – Berkeley, US)

Joint work of Edward A. Lee, Erling Jellum

This talk will introduce the Distributed Rhythm Generator, a distributed real-time program written in Lingua Franca. Distributed real-time systems are intriguingly challenging systems, as they must balance two opposing requirements. First, they must often only actuate based on a consistent view of the shared variables of the overall system, and second, they must actuate in a timely manner. Without time-predictable networks, reconciling these requirements might not be possible and a tradeoff must be made. We will show how the reactor-oriented programming paradigm introduced by Lingua Franca, enables an elegant specification of such consistency-available tradeoffs.

3.14 Time-Tarot: Toward a Quantitative Approach to Time-Predictability

Chadlia Jerad (University of Manouba, TN)

Joint work of Chadlia Jerad, Martin Schoeberl, Emad Jacob Maroun, Edward A. Lee, Shaokai Lin, Erling Jellum

The current trends in hardware design aim at improving the average performance. Together with the increase in software complexity, this trajectory runs contrary to time-predictability, especially when designing safety-critical and real-time systems, as timing requirements need to be guaranteed a priori. Literature is rich with valuable attempts to define and measure time-predictability. Still, a quantitative approach that covers all the different features at the different abstraction levels is missing. This talk suggests the design of a concept map for time-predictability levels. It is based on results from different research papers and is meant to be extensible to accommodate future inventions. It is also derived from expert knowledge in the different fields. The concept map, when combined with multi-criteria decision techniques, can formalize the choice among two or more designs. The techniques to be used are the Brown-Gibson Model and the Analytic Hierarchy Process. The former has the advantage of combining objective and subjective factors that immediately derive from the concept map. Because weighted scoring is required in this method, we use the Analytic Hierarchy Process to weigh the features among each other. A challenge though is about the support of multi-processor systems with different architectures. In addition, the use of the worst-case performance can be decided to be used either as a validation tool or to perform fusion of the results.

3.15 Declarative Lifecycle Management in Digital Twins

Einar Broch Johnsen (University of Oslo, NO)

Joint work of Einar Broch Johnsen, Eduard Kamburjan, Nelly Bencomo, Silvia Lizeth Tapia Tarifa

Together, a digital twin and its physical counterpart can be seen as a self-adaptive system: the digital twin monitors the physical system, updates its own internal model of the physical system, and adjusts the physical system by means of controllers in order to maintain given requirements. As the physical system shifts between different stages in its lifecycle, these requirements, as well as the associated analyzers and controllers, may need to change. The exact triggers for such shifts in a physical system are often hard to predict, as they may be difficult to describe or even unknown; however, they can generally be observed once they have occurred, in terms of changes in the system behavior. This talk presents an automated method for self-adaptation in digital twins to address shifts between lifecycle stages in a physical system, based on recent work [1]. Our method is based on declarative descriptions of lifecycle stages for different physical assets and their associated digital twin components. Declarative lifecycle management provides a high-level, flexible method for self-adaptation of the digital twin to reflect disruptive shifts between stages in a physical system.

References

1 Eduard Kamburjan, Nelly Bencomo, Silvia Lizeth Tapia Tarifa, and Einar Broch Johnsen. Declarative lifecycle management in digital twins. In *Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems*, MODELS Companion '24, page 353–363, New York, NY, USA, 2024. Association for Computing Machinery.

3.16 Security vs. Accessibility Tradeoffs in Reactive Systems

Hokeun Kim (Arizona State University – Tempe, US)

In this framing discussion, I explore the fundamental tradeoffs between security and accessibility in designing reactive and cyber-physical systems (CPS). This discussion examines how secure access mechanisms – such as authentication, encryption, detection, and mitigation – must balance with the need for accessibility. Through real-world examples and case studies, I highlight the challenges of designing secure reactive systems and CPS while managing the tradeoffs between security and accessibility, considering threat models, costs, and security requirements of the target systems. Various design decisions, including connectivity, encryption strategies, and the choice between prevention and detection of cyber threats, are analyzed in the context of modern threat models. This framing discussion also presents a methodology for enforcing security without sacrificing essential accessibility, ensuring the domain-specific reactive systems and CPS are secure and protected yet accessible.

3.17 Consistency vs. Availability in Reactive Systems Design

Edward A. Lee (University of California – Berkeley, US)

Joint work of Edward A. Lee, Ravi Akella, Soroush Bateni, Shaokai Lin, Marten Lohstroh, Christian Menard

Distributed software systems often require consistent shared information. For example, connected vehicles require agreement on access to an intersection before entering the intersection. It is far from trivial, however, how to achieve consistency, or even how to define it rigorously enough to know when it has been achieved. In this talk, I will show how strong and weak forms of consistency can be defined, how software infrastructure can provide reasonable guarantees and efficient implementations, and what are the fundamental costs of achieving consistency that no software system can avoid. Specifically, I will outline the CAL theorem, which quantifies consistency, availability, and latency, and gives an algebraic relation that shows that as latency increases, either availability or consistency or both must decrease. I will describe a coordination language called Lingua Franca that enables programmers to explicitly work with the tradeoffs between these three quantities.

3.18 CAL Theorem in Reactive Systems

Shaokai Jerry Lin (University of California – Berkeley, US)

Joint work of Shaokai Jerry Lin, Edward A. Lee

In 2000, Eric Brewer (UC Berkeley) famously stated the CAP theorem during the PODC keynote. However, the tradeoffs mentioned in the theorem was not formally specified. Lee et al. [1] later proposed the CAL theorem, which establishes a mathematical relationship among the three properties: Consistency, Availability, and Latency (extending Partitioning in the CAP theorem). In this talk, we present a work-in-progress formalism based on the CAL theorem, which aims to 1) be easily applicable to programming frameworks, and 2) capture the behavior of a reactive system over time. We demonstrate the formalism on a simple stock exchange application written in the Lingua Franca (LF) coordination language. We derive timing constraints from the semantics of the LF language and the application, formulating a system of equations calculating earliest firing times of reactions from previous firing times and asynchronous inputs. We represent the system of equations into a linear difference equation using the max-plus algebra and show that the eigenvalue of a coefficient matrix represents the threshold of the system's responsiveness.

References

1 Edward A. Lee, Ravi Akella, Soroush Bateni, Shaokai Lin, Marten Lohstroh, and Christian Menard. Consistency vs. availability in distributed cyber-physical systems. *ACM Trans. Embed. Comput. Syst.*, 22(5s), September 2023.

3.19 Benchmarking Worst-Case Performance of Real-Time Systems

Martin Schoeberl (Technical University of Denmark – Lyngby, DK), Erling Jellum (University of California – Berkeley, US), Shaokai Lin (University of California – Berkeley, US), Chadlia Jerad (Univrsity of manouba, TN), Emad Jacob Maroun (Technical University of Denmark – Lyngby, DK), Marten Lohstroh (University of California – Berkeley, US), and Edward A. Lee (University of California – Berkeley, US)

Real-time systems rely on tasks with well-defined worst-case execution times (WCET) to ensure predictable behavior. To facilitate accurate WCET estimation, some researchers advocate for processor architectures that simplify timing analysis, known as precision-timed or time-predictable architectures. However, the precise definition and quantification of time predictability remain open questions. This talk examines the concept of time predictability and the challenges in measuring it. Instead of focusing solely on architectural properties, we argue that worst-case performance should be evaluated as a combined property of the processor, compiler, and WCET analysis tool. To enable systematic evaluation, we propose to standardize on a benchmark suite for assessing time-predictable processors, compilers, and WCET analysis tools. We define worst-case performance as the geometric mean of WCET bounds across this benchmark set.

3.20 Certification of ML-based systems

Claire Pagetti (ONERA – Toulouse, FR) and Arthur Clavière (Collins Aerospace – Blagnac, FR)

The purpose of the talks was to present some of the results reached for introducing ML (Machine Learning) algorithms in Airborne systems. First a brief introduction on certification and on current drafted standards for ML was done. Indeed, EASA is writing roadmaps and preliminary guidance. SAE/EUROCAE working group is in parallel writing a future ED/ARP document. The presentation then has dug to two use cases. The first was the simplest: how to replace a set of LUT (look-up tables) – here for the avoidance collision system ACAS Xu – with neural networks while preserving the intended function. For that, formal verification was successfully applied. These kinds of surrogate models to compress existing code is a way forward. The seconds case is exploratory: vision-based perception algorithm to detect runway during an aircraft landing. Again formal verification was successfully applied to check robustness against some foreseeable perturbations. These preliminary results open the path for certification.

3.21 Synergies between Timing Predictability and Microarchitectural Security

Jan Reineke (Universität des Saarlandes – Saarbrücken, DE)

Joint work of Marco Guarnieri, Gideon Mohr, Zilong Wang, Klaus v. Gleissenthall, Jan Reineke

Timing attacks exploit variations in execution time to leak sensitive information. Hardware-software leakage contracts are a new security abstraction that augments the instruction-set-architecture (ISA) to capture microarchitectural leakage at software level. In recent work we have shown how to verify that open-source RISC-V processors satisfy particular contracts and how to synthesize precise leakage contracts for a given processor. In this talk, I motivate and introduce leakage contracts and I discuss two equivalent notions of contract satisfaction that suggest synergies between research on timing predictability and microarchitectural security.

3.22 Observations on Formalizing Reactive Systems

Marcus Rossel (Barkhausen Institut – Dresden, DE)

Reactive systems have a rich history of rigorous modeling and verification techniques. As modern verification efforts become more ambitious, they require the composition of a growing number of techniques. This composition necessitates interfacing between different formalisms and tools, which can lead to error-prone discontinuities in the verification pipeline. While there exist systems for coherently verifying reactive systems, they are limited in expressivity. Interactive theorem provers (ITPs) provide a means of verification with virtually unbounded expressiveness, but do not natively capture the semantics of notions like concurrency and mutability. We discuss the problems encountered when attempting gap-free verification of reactive systems in ITPs.

3.23 Timing Tradeoffs in Timed Automata with Dynamic Ticks

Alexander Schulz-Rosengarten (Universität Kiel, DE) and Reinhard von Hanxleden (Universität Kiel, DE)

Synchronous languages build on an abstraction from physical execution time by dividing the execution into logical ticks. However, they say little about when to execute the ticks and traditionally lack built-in support for physical time. This makes it rather cumbersome to express things like time-outs or periodic executions. We present Timed SCCharts that use a timed automata formalism and combine it with the concept of dynamic ticks. This enables specifying time-dependent behavior while addressing tradeoffs in timing in the face of imperfections of execution with physical time.

One application area we plan to explore is the railway domain, as part of the REAKT initiative. This initiative aims to revitalize rural rail lines, using the line Bad Malente – Lütjenburg as real world laboratory. As an aside, in Germany alone, about 5000 km of rural train lines have been put out of service since the 1990s. This corresponds to about 30% of the total network, and it directly affects more than 3 million citizens.

3.24 Fail-operational Systems in Autonomous Driving Applications

Katharina Sedow (Saneon GmbH – Ismaning, DE)

As the automotive research and development field approaches automated driving levels SAE L4 and L5, the introduction of the requirement for the driving function to continue operation when an abnormal system behavior occurs becomes crucial. Highly automated and autonomous systems cannot use a human driver as the fallback decision-maker and actuator, and, in order to be safe, they must possess the ability to function until the vehicle reaches a safe state. This talk briefly introduces the concept behind fail-operational systems. First, we introduce the advances of such systems when compared to the fail-safe implementations. We further provide an example of a fail-operational architecture. In the last part of the talk, we discuss the challenges the automotive industry faces when developing fail-operational driving functionality. They include the assurance of independent redundancy, tradeoffs in performance and safety, as well as the development costs. Finally, we address the approaches intended to handle those issues, such as redundant system design, the introduction of monitoring safety components, and the handling of the high safety integrity levels.

3.25 The Functional, the Imperative, and the Sudoku

Manuel Serrano (INRIA – Sophia Antipolis, FR)

Joint work of Manuel Serrano, Robert Bruce Findler

Conventional wisdom suggests that the benefits of functional programming no longer apply in the presence of even a small amount of imperative code, as if the addition of imperative code effectively subtracts. And yet, as we show in this talk, combining functional programming with the special imperative language Esterel provides a multiplicative improvement to the benefits of functional programming.

To illustrate these benefits, the bulk of this talk consists of an in-depth exploration of HipHop code (a mashup of JavaScript and Esterel) that implements a Sudoku solver, showing how it is possible to write code that is as easy to understand as if it were written in a pure functional programming style, even though it uses multiple threads, mutable state, thread preemption, and even thread abortion. Even better, concurrent composition and task canceling provide significant program structuring benefits that allow a clean decomposition and task separation in the solver.

3.26 Consistency versus Availability in Redundant Controllers – Formal Verification, Test Design, Time Analysis

Marjan Sirjani (Mälardalen University – Västerås, SE)

A potential problem that may arise in the domain of distributed control systems is the existence of more than one primary controller in redundancy plans which may lead to inconsistency. We worked on an algorithm called NRP FD (Network Reference Point Failure Detection), proposed by industry, to solve this issue by prioritizing consistency over availability. I explain how by using modeling and formal verification, we discovered an issue in NRP FD where we may have two primary controllers at the same time. We then provide a solution to mitigate the identified issue, thereby enhancing the robustness and reliability of such systems. In the same context, I will also show how we used model checking for making informed decisions in designing test cases for fault-tolerant systems. I will add a discussion on how this approach may be generalized in different contexts.

3.27 Tradeoffs in Accuracy and Timeliness in Transportation Cyber-Physical Systems

Jonathan Sprinkle (Vanderbilt University – Nashville, US)

Joint work of Jonathan Sprinkle, George Gunter, Daniel B. Work, the CIRCLES Consortium

This talk explores the timing and accuracy tradeoffs from application problems in transportation cyber-physical systems. The talk demonstrates previous proof of dampening traffic waves in closed road scenarios, and evidence that existing adaptive cruise controllers amplify (rather than reduce) traffic waves. To address these issues, the work discusses the use of hierarchical high-latency and low-latency controllers, which were shown to be successful. These approaches utilize shared information of the road network that may be gathered over recent minutes, for the purposes of reducing stop and go traffic waves. Given the dynamics of freeway traffic, where speeds of up to 35 m/s regularly encounter stop and go waves traveling at -4 m/s (i.e., against the flow), the lookahead of the on-vehicle sensors is insufficient to dampen these waves, and external information is needed. Using only local sensors it is possible to avoid collisions, but it is not trivial to also dampen the traffic waves. Thus it is important to understand how to share data on downstream state information, with timing information to help understand how to interpret those data. The talk transitions to discussion on how to design and utilize safety envelope controllers for future experiments, in which AI or other untrusted controllers are fielded in open road scenarios. While these controllers can be theoretically shown to be safe, the theory requires information from vehicle sensors (such as derivatives of velocity or acceleration) which are difficult to obtain, and for which filtering delays the availability of the data – requiring an additional following gap for safety that may reduce the effectiveness of the approach. The conclusion calls for continued exploration in determining the accuracy of these distributed systems, which are loosely coordinated, when it is impossible to recreate the situations in which they were fielded. Further, it calls for understanding how one can safely field AI-based controllers to gather field training data, when there may be significant limitations in understanding how safe the safety envelopes are for general use.

This work is supported by the US National Science Foundation, and the US Department of Energy.

3.28 Timeliness vs. Accuracy

Lothar Thiele (ETH Zürich, CH)

This talk surveys the research landscape surrounding the fundamental tradeoff between timeliness and accuracy in complex systems. It explores formal and informal approaches to managing this tradeoff from multiple perspectives, including neuroscience and psychology, scheduling theory, algorithm design, system architecture, and multi-agent systems. Connecting research from the late 1980s to cutting-edge developments, the talk examines how current abstractions, particularly those related to accuracy (e.g., algorithmic precision) and timeliness (e.g., system timing requirements), hold up in the face of increasing system complexity. A core question explored is whether these abstractions remain adequate, or if we need fundamentally new approaches beyond reliance on testing and benchmarking in real-world environments, to effectively address the timeliness-accuracy tradeoff across algorithms, runtime systems (including hardware), and requirements. Finally, the talk identifies key open research questions in this area.

3.29 Fundamental Tradeoffs in Reactive Systems for Smart Agriculture and Pollutants Detection in Resource-Constrained Environments

Eric Tutu Tchao (Kwame Nkrumah University of Science and Technology, GH)

In resource-constrained environments like rural Ghana, designing reactive systems for critical applications, such as smart agriculture and detecting pollutants from illegal mining activities, demands confronting unavoidable tradeoffs. These systems must balance competing priorities shaped by limited energy, connectivity, and financial resources. This has forced researchers to examine how fundamental engineering limits force compromises between accuracy, timeliness, security, and scalability. Researchers cannot eliminate compromises but formalize them, embedding transparency into system architecture. By doing so, they create solutions that are survivable, equitable, and aligned with the pragmatic realities of communities like Ghana's – where "good enough" is not a concession but a necessity. My presentation at Dagstuhl Seminar 25091 illustrates a broader truth that in resource-constrained environments, tradeoffs are not design failures but fundamental constraints.

3.30 Utility-Based System Design: Making Sense of Tradeoffs

Eugene Yip (GLIWA – Weilheim, DE)

Joint work of Eugene Yip, Gerald Lüttgen

Selecting software components to reuse as a means to reduce development time and costs can be challenging, especially when multiple components provide the same functionality but offer different non-functional properties, e.g., numerical accuracy, memory footprint, power rating, timeliness, robustness, security, or financial cost. In this talk, we seek to model the benefit that a system gains from a component's non-functional properties using so-called "Property-Utility Functions" (PUFs) as a means to formally analyse tradeoffs in component selection. We will discuss how a PUF can model the incremental benefit (utility) that a system would receive for each additional unit of resource that a component consumes. We will explore open challenges in how the utility of a datum or event is transformed by the PUFs of the components they flow through, how to tradeoff one component for another based on their qualities, how to compose and decompose PUFs for bottom-up and top-down system design, and how to diagnose violations in expected utilities.

4 Working groups

4.1 Can AI be used in critical Systems?

Jérôme Hugues (Carnegie Mellon University – Pittsburgh, US), Andres Barrilado (NXP Semiconductors – Toulouse, FR), Frédéric Boniol (ONERA – Toulouse, FR), Thomas Carle (Toulouse University, FR), Jerónimo Castrillón-Mazo (TU Dresden, DE), Samarjit Chakraborty (University of North Carolina at Chapel Hill, US), Arthur Clavière (Collins Aerospace – Blagnac, FR), Victor Jegu (Airbus S.A.S. – Toulouse, FR), Claire Pagetti (ONERA – Toulouse, FR), Marcus Rossel (Barkhausen Institut – Dresden, DE), Selma Saidi (TU Braunschweig, DE), Jonathan Sprinkle (Vanderbilt University – Nashville, US), Hasna Bouraoui (TU Dresden, DE) and Eugene Yip (GLIWA – Weilheim, DE)

The working group concluded that deploying AI in safety-critical systems hinges on balancing two tightly linked factors – accuracy and timeliness – because longer training or inference windows can enhance correctness but may violate real-time constraints. Across the four domains discussed (avionics, automotive, medical devices such as pacemakers, and rail), regulators are cautiously expanding standards: avionics currently only certifies off-line–trained ML components, automotive guidelines are rapidly evolving, medical devices show promising results under tight safeguards, and rail is changing slowly. Participants agreed AI is most defensible when it supplements rather than replaces human judgment or expensive hardware (e.g., monocular camera + AI distance estimation versus LiDAR) and when robust mitigation layers – observers, simplex architectures, safety "nets," or treating AI like an imperfect sensor – contain the impact of inevitable errors. Defining what constitutes a "good" AI therefore means specifying clear functional and safety requirements, acknowledging residual failure probabilities, and ensuring retraining or adaptation does not void certification credits. Finally, the group underscored that verification strategies must scale with AI complexity: small-state reinforcement learning can leverage formal methods, whereas high-dimensional perception models require new notions of coverage, traceability, and explainability, reinforcing that full autonomy remains out of reach today while incremental, supervised uses continue to gain traction.

4.2 Distributed music challenge

Grzegorz Bazydło (University of Zielona Gora, PL), Anupam Chattopadhyay (Nanyang TU – Singapore, SG), Andrés Goens Jokisch (University of Amsterdam, NL), Chadlia Jerad (University of Manouba, TN), Erling Rennemo Jellum (University of California – Berkeley, US), Einar Broch Johnsen (University of Oslo, NO), Hokeun Kim (Arizona State University – Tempe, US), Edward A. Lee (University of California – Berkeley, US), Shaokai Jerry Lin (University of California – Berkeley, US), Jan Reineke (Universität des Saarlandes – Saarbrücken, DE), Manuel Serrano (INRIA – Sophia Antipolis, FR), Martin Schoeberl (Technical University of Denmark – Lyngby, DK), Lothar Thiele (ETH Zürich, CH) and Reinhard von Hanxleden (Universität Kiel, DE)

In the breakout group session titled "Distributed Music Challenge," we explored the practical aspects of the Distributed Rhythm Generator introduced in the preceding talk. This hands-on activity involved running a distributed real-time program written in Lingua Franca across two laptops, simulating the behavior of a networked rhythm generator. The session provided a concrete illustration of the tradeoffs between consistency and availability in distributed real-time systems. Participants experimented with three distinct conditions: normal network operation, a scenario with network overload, and one with disrupted clock synchronization. Through these scenarios, the group observed how the Lingua Franca runtime handled timing and consistency under different constraints, deepening the discussion on how reactor-oriented programming can be used to model and manage such tradeoffs elegantly in distributed systems.

4.3 Orchestration/coordination languages vs reactive languages

Jérôme Hugues (Carnegie Mellon University – Pittsburgh, US), Edward A. Lee (University of California – Berkeley, US), Shaokai Jerry Lin (University of California – Berkeley, US), and Manuel Serrano (INRIA – Sophia Antipolis, FR)

The breakout group discussed the relationship between coordination languages, synchronous reactive languages, and choreographic languages. First, we started by reviewing examples of each language category, namely Reo, Choral, Rebeca, Esterel, and Lingua Franca. The group recognized that all these languages do some form of coordination. Some also have the capability to support the full implementation either within their own language such as Esterel, or through the integration with other programming languages like Lingua Franca. In other cases, the functional code is considered opaque, and the language only defines the interface of these containers.

The group noted that these languages define how the synchronization of the execution of some containers happen. More specifically, their semantics define when (time), why (causality), and which action is executed. These containers are stateful, and they react to multiple, possibly simultaneous, inputs. Hence, it is critical to address when these events are sent, how they are processed, and how to change the container's state to guarantee determinism. These questions become more complicated when considering an implementation on resource-constrained embedded targets, as opposed to larger machines.

Finally, the group noted that this reasoning can be extended to not just mathematical formalisms such as communicating sequential processes (CSP) or Kahn process networks (KPN), but also synchronization/communication APIs such as POSIX or MPI. They also define some notion of stateful containers. However, in the case of API-based solutions, the actual boundary of the container and its interface is not a first-class citizen: asserting the correctness or determinism of a concurrent system, e.g., built on top of POSIX APIs, is more complex than a similar system expressed in a dedicated coordination language.

4.4 Tradeoffs in accuracy and timeliness

Andres Barrilado (NXP Semiconductors – Toulouse, FR), Hasna Bouraoui (TU Dresden, DE), Anupam Chattopadhyay (Nanyang TU – Singapore, SG), Jerónimo Castrillón-Mazo (TU Dresden, DE), Arpan Gujarati (University of British Columbia – Vancouver, CA), Hokeun Kim (Arizona State University – Tempe, US), Lothar Thiele (ETH Zürich, CH), Alexander Schulz-Rosengarten (Universität Kiel, DE), Katharina Sedow (Saneon GmbH – Ismaning, DE) and Jonathan Sprinkle (Vanderbilt University – Nashville, US)

This breakout session addressed the challenges, methodologies, and future directions involved in managing tradeoffs between accuracy and timeliness in computational systems. The central theme built on Lothar Thiele's earlier seminar presentation and drew parallels with Thinking, Fast and Slow [1], asking how systems could dynamically respond with appropriately timed decisions and varying confidence levels. The key questions and themes were:

- What methodologies can be used to manage accuracy vs. timeliness tradeoffs?
- What are the real costs of missing a timeline in industry today?
- How do real-time systems currently handle uncertainty or degraded performance?
- What is the role of resource scaling (e.g., parallelism) in trading accuracy for time?

During the discussion, Andrès Goens emphasized the benefits of early-exit strategies and iterative answer refinement for real-time constrained systems, raising concerns about how uncertainty compounds in extended predictions – illustrated with a motor motion example where the sine wave frequency is unknown. He suggested connecting these ideas to computationally aware model predictive control (MPC), citing Jonathan's earlier work, and posed the question: "Getting something quick converges eventually, but what issues expand with extended predictions?". Anupam Chattopadhyay contributed by stressing the importance of leveraging parallel resources to improve accuracy under time constraints, especially in embarrassingly parallel applications. He framed timeliness and accuracy as just two among many competing objectives, and discussed the challenge of managing out-of-order answers in anytime systems, noting that "getting more accuracy with unbounded time might not be worth it."

The discussion on use case and domain examples highlighted challenges in automotive and rail systems, particularly regarding decision timelines, fallback mechanisms, and uncertainty handling. In autonomous vehicles across L2 to L5 levels, issues such as classification errors – like braking for an image of a person rather than a real one – underscored the difficulty of making safe decisions when initial data is ambiguous. Real-world cases included Waymo's geofenced operations in Phoenix and the contrast between fast emergency fallback systems and more complex uncertainty mitigation strategies in advanced vehicles. In rail systems, the focus shifted to multi-modal data fusion from LiDAR and cameras, emphasizing the challenge of resolving inconsistencies, especially when distinguishing between humans and other objects on the tracks. The layered processing approach – comprising rapid object detection, slower identification, and subsequent motion tracking – was presented as a method to incrementally refine understanding while balancing timeliness and accuracy.

Human-centered analogies highlighted how people naturally adjust processing time based on perceived threat and proximity, serving as an implicit model for soft real-time systems. For example, humans allocate more cognitive resources to assessment when a potential hazard appears distant, whereas immediate threats like the need to brake demand rapid reaction. Additionally, the way distractions – such as animals – can temporarily obscure more critical upcoming stimuli like vehicles illustrates how prioritization and attention shifts occur dynamically, a concept relevant to designing systems that must manage changing real-time demands.

Key technical considerations included debates on online training and distribution drift, particularly whether it is preferable to adapt models dynamically at runtime or to preserve memory of rare edge cases that could be critical. Shadow mode simulation emerged as a strategy for evaluating new algorithms alongside active systems, especially within virtual or "metaverse" environments, offering a non-intrusive way to test improvements. Additionally, sensor cascades were discussed, where secondary sensors are activated based on outputs from primary sensors to refine decision accuracy, enabling more responsive and adaptive perception pipelines.

To visually support and extend the conceptual discussions, several figures (Figure 1, Figure 2, Figure 3) were sketched throughout the session. These illustrations capture the tradeoffs between accuracy, time, and computational effort in real-time and soft real-time systems. They also highlight architectures that mirror human-inspired decision strategies and progressive refinement models.

A key focus of the discussion was on the complex challenges of designing systems that must operate across multiple timelines while adapting to shifting, context-sensitive deadlines. Participants explored how systems can meaningfully react to outdated but potentially valuable information, and the risks associated with making premature decisions based on incomplete data. A recurring theme was the importance of metadata – such as timestamps and confidence scores – to inform decision-making under uncertainty. Additionally, the need to establish acceptance criteria for degraded or "poisoned" data, and the integration of error-correcting mechanisms in fault-tolerant designs, was emphasized. These challenges raised several open questions about the future of dynamic system design.

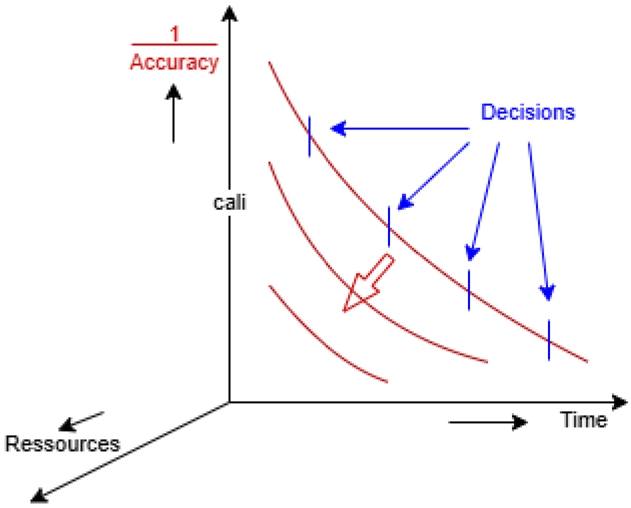

Figure 1 Anytime Algorithm Convergence: Anytime algorithms will converge on a specific accuracy, which improves over time. Understanding any mid-execution accuracy (decision points) can help approximate what the final accuracy may be. Additional resources for parallel activity may further improve the convergence time. Axes: Y-axis: accuracy, X-axis: time, Z-axis: resources required to obtain the particular decision points).

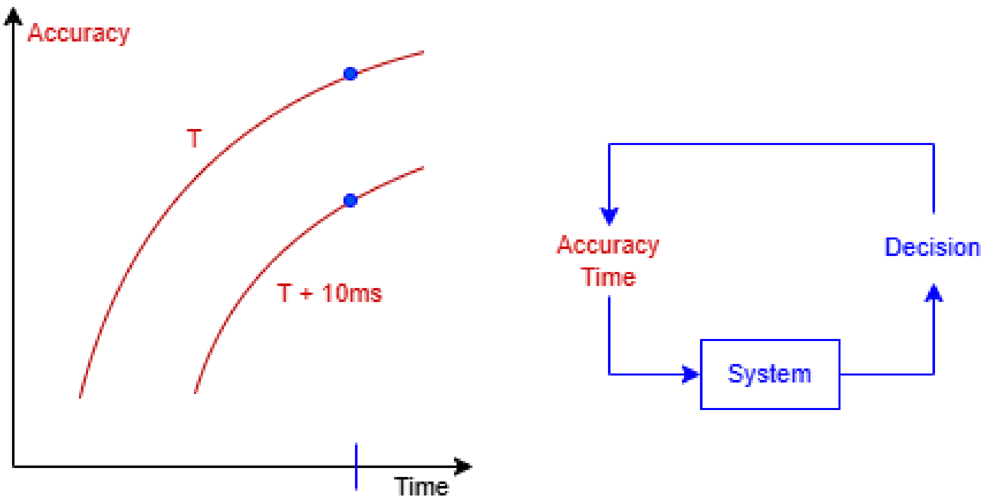

Figure 2 Frame-based Decision Enhancement: Taking into account regular processing of frame rates: enabling parallel access with in-process updates (not early exit) could open the pathway to taking 2 data values from two frames, and using the combined information to make a more informed decision (Axes: Y-axis: accuracy, X-axis: time).

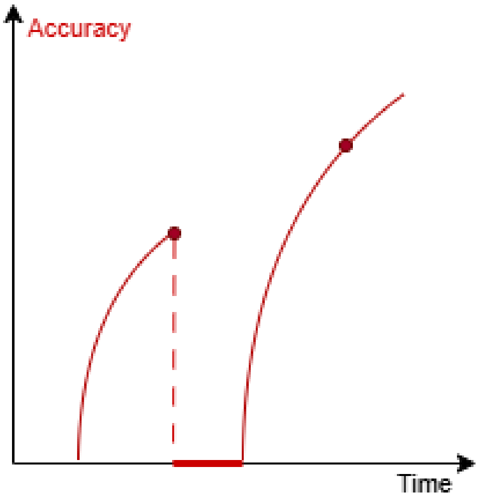

Figure 3 A "Thinking Fast, Thinking slow" explicit example: Component 1 can arrive with something above the accuracy threshold, but will not maximize accuracy quickly. The result from its quick response can trigger the next component, which should (with initial criteria) arrive at a better solution, faster than starting without that information. Axes: Y-axis: accuracy, X-axis: time).

Key System Design Challenges:
- Managing multiple, context-sensitive timelines and deadlines
- Handling outdated but potentially valuable data
- Mitigating the cost of incorrect early decisions
- Timestamping decisions and including confidence metadata
- Defining acceptance criteria for incomplete or degraded input
- Building fault-tolerant systems that can still reason under uncertainty

Open Questions:
- How can contextual deadlines be defined in reactive systems?
- What are effective methods to systematically integrate fault models and fallback strategies?
- Beyond automotive and rail, which domains are best suited for designs emphasizing dynamic tradeoffs in accuracy and timeliness?

As a closing summary, the group emphasized that:
- Timeliness vs. Accuracy must be treated not as static tradeoffs but as part of a system's runtime behavior.
- Current systems rarely pay a high cost for delayed decisions, but future high-stakes autonomous platforms (e.g., L5 vehicles, medical robotics) will.
- There is opportunity in revisiting layered decision systems, sensor fusion under uncertainty, and the reuse of older data in dynamic ways.
- Jerónimo Castrillón-Mazo's Debriefing Thought: "When the trigger changes your deadline, you are in a new kind of reactive system."

References

1 Daniel Kahneman. *Thinking, fast and slow*. macmillan, 2011.

4.5 Benchmarks for real-time systems

Thomas Carle (Toulouse University, FR), Samarjit Chakraborty (The University of North Carolina (UNC) at Chapel Hill, US), Arthur Clavière (Collins Aerospace – Blagnac, FR), Victor Jegu (Airbus S.A.S. – Toulouse, FR), Chadlia Jerad (University of Manouba, TN), Claire Pagetti (ONERA – Toulouse, FR), Jan Reineke (Universität des Saarlandes – Saarbrücken, DE) and Martin Schoeberl (Technical University of Denmark – Lyngby, DK)

The work group on "Benchmarks for Real-Time Systems" discussed the shortcomings of existing benchmark suites like TACLeBench and explored possible directions for developing more representative and useful benchmarks. TACLeBench was criticized for being largely composed of small, non-real-time benchmarks, lacking multi-threaded code, and using only fixed inputs. Furthermore, it does not offer standardized licensing, nor does it provide clear guidance on the context in which benchmarks should be executed or evaluated. These limitations make it ill-suited for evaluating modern real-time systems, which are increasingly complex and heterogeneous.

Participants noted that existing real-time benchmarks are used for a range of evaluation purposes – such as analyzing the worst-caes performance and/or predictability of microarchitectures, assessing WCET (Worst-Case Execution Time) analysis tools, and evaluating scheduling algorithms. Depending on the benchmarks' purpose they could consist of individual tasks, full applications, or even complete systems where specific software and hardware combinations are considered together.

A key theme in the discussion was the importance of designing benchmarks that reflect actual real-time workloads and applications. Real-time systems span diverse domains and benchmarks should reflect this diversity. Instead of relying exclusively on small code fragments, participants suggested a potential shift towards specification-based benchmarks: defining the intended behavior rather than a fixed implementation, with the goal of deriving implementations that exhibit predictable worst-case performance.

Domains that are particularly representative of modern real-time systems include vision and perception (e.g., camera or LIDAR input), sensor fusion, control algorithms such as model-predictive control, and mixed workloads running on heterogeneous architectures (e.g., involving both CPUs and GPUs). These domains may feature execution time variation depending on the nature of the input data, a characteristic that existing benchmark suites largely fail to capture.

Several participants pointed to tools like Simulink and Lustre as valuable sources for generating realistic benchmarks. Simulink, in particular, includes many example controllers that are already central to real-world safety-critical systems. These models can be used to generate C code, with support for different code generation configurations, including full closed-loop systems that simulate both controllers and plants. Similarly, code generated from Lustre, or real-world code bases like Autoware, could serve as useful starting points for building a more diverse and representative benchmark suite.

Another important aspect raised was the need to define reference contexts for benchmarks. This includes specifying the hardware architecture and microarchitecture, selecting reasonable parameter values, and ideally providing baseline results obtained using known analysis tools. Benchmarks should also include annotations – such as loop bounds – to ensure they are compatible with WCET analysis tools like AbsInt.

5 Software teasers

In this session, various seminar participants demonstrate programming languages and tools they are developing.

5.1 Timed Rebeca

Marjan Sirjani (Mälardalen University – Västerås, SE)

Timed Rebeca is an actor-based modeling language equipped with a model-checking tool. It is designed to be both accessible and usable for software engineers and grounded in formal semantics, ensuring formal verification support. A cyber-physical system is represented as a set of communicating actors, and safety and liveness properties can be verified using the model-checking tool, Afra. Rebeca homepage can be accessed here: `https://rebeca-lang.org/`.

5.2 Lingua Franca

Erling Rennemo Jellum (University of California – Berkeley, US), Edward A. Lee (University of California – Berkeley, US)

Lingua Franca is a reactor-oriented coordination language for implementing cyber-physical systems. Due to its declarative syntax, automatic diagramming is possible. By separating the concerns of coordination and timing from the computation, it is a promising target for code generated by large language models. Lingua Franca homepage can be accessed here: `https://www.lf-lang.org/`.

5.3 SCCharts

Alexander Schulz-Rosengarten (Universität Kiel, DE)

SCCharts is a reactive synchronous programming language with sequentially constructive semantics. It supports dynamic ticks, which are driven by a minimal interface to the environment. As such, it may be integrated with other runtimes, such as Lingua Franca, to provide its ticks. SCCharts rely on several model-to-model transformations to turn an SCChart model into code. SCCharts has sophisticated diagram generation, the same that is used in Lingua Franca. Each compilation step can also be inspected graphically, making the compiler transparent and easing development. Links to SCCharts can be accesses here: `https://github.com/kieler/semantics/wiki/Quick-Overview`.

5.4 QRML

Marc Geilen (TU Eindhoven, NL)

QRML is a domain-specific language for formulating multi-objective optimization problems for cyber-physical systems. The problems can be exported to several formats and solved by tools like Z3. QRML also runs in the browser. QRML tool can be accessed here: `https://qrml.org/qrml`.

5.5 HipHop

Manuel Serrano (INRIA – Sophia Antipolis, FR)

HipHop is a synchronous-reactive language embedded in Javascript. It connects the asynchronous world of Javascript with the synchronous world of Esterel. The HipHop program describes a deterministic and concurrent response to external inputs, and the embedding Javascript decides when to trigger the HipHop program. HipHop can be downloaded from GitHub (`https://github.com/manuel-serrano/hiphop`).

Participants

- Andres Barrilado
NXP Semiconductors – Toulouse, FR
- Grzegorz Bazydlo
University of Zielona Gora, PL
- Frédéric Boniol
ONERA – Toulouse, FR
- Hasna Bouraoui
TU Dresden, DE
- Thomas Carle
Toulouse University, FR
- Jerónimo Castrillón-Mazo
TU Dresden, DE
- Samarjit Chakraborty
University of North Carolina at Chapel Hill, US
- Anupam Chattopadhyay
Nanyang TU – Singapore, SG
- Arthur Clavière
Collins Aerospace – Blagnac, FR
- Marc Geilen
TU Eindhoven, NL
- Alain Girault
INRIA – Grenoble, FR
- Andrés Goens Jokisch
University of Amsterdam, NL
- Arpan Gujarati
University of British Columbia – Vancouver, CA
- Jérôme Hugues
Carnegie Mellon University – Pittsburgh, US
- Victor Jegu
Airbus S.A.S. – Toulouse, FR
- Erling Rennemo Jellum
University of California – Berkeley, US
- Chadlia Jerad
University of Manouba, TN
- Einar Broch Johnsen
University of Oslo, NO
- Hokeun Kim
Arizona State University – Tempe, US
- Edward A. Lee
University of California – Berkeley, US
- Shaokai Jerry Lin
University of California – Berkeley, US
- Claire Pagetti
ONERA – Toulouse, FR
- Jan Reineke
Universität des Saarlandes – Saarbrücken, DE
- Marcus Rossel
Barkhausen Institut – Dresden, DE
- Selma Saidi
TU Braunschweig, DE
- Klaus Schneider
RPTU Kaiserslautern-Landau, DE
- Martin Schoeberl
Technical University of Denmark – Lyngby, DK
- Alexander Schulz-Rosengarten
Universität Kiel, DE
- Katharina Sedow
Saneon GmbH – Ismaning, DE
- Manuel Serrano
INRIA – Sophia Antipolis, FR
- Marjan Sirjani
Mälardalen University – Västerås, SE
- Jonathan Sprinkle
Vanderbilt University – Nashville, US
- Eric Tutu Tchao
Kwame Nkrumah University of Science and Technology, GH
- Lothar Thiele
ETH Zürich, CH
- Reinhard von Hanxleden
Universität Kiel, DE
- Eugene Yip
GLIWA – Weilheim, DE

Report from Dagstuhl Seminar 25092

Estimation-of-Distribution Algorithms: Theory and Applications

Josu Ceberio Uribe[*1], **Benjamin Doerr**[*2], **John McCall**[*3], **Carsten Witt**[*4], and **Marcus Schmidbauer**[†5]

1 **University of the Basque Country – Donostia, ES.** josu.ceberio@ehu.eus
2 **Ecole Polytechnique – Palaiseau, FR.** doerr@lix.polytechnique.fr
3 **The Robert Gordon University – Aberdeen, GB.** j.mccall@rgu.ac.uk
4 **Technical University of Denmark – Lyngby, DK.** cawi@dtu.dk
5 **Universität Passau, DE.** marcus.schmidbauer@uni-passau.de

Abstract

The Dagstuhl Seminar 25092 “Estimation-of-Distribution Algorithms: Theory and Practice” took place on February 23–28, 2025. It brought together 25 international experts in estimation-of-distribution algorithms (EDAs) and related research areas. Their research focus ranged from theoretical aspects like mathematical runtime analyses to efficient solutions of real-world problems in industrial contexts. This report documents the program and the main outcomes of this fruitful seminar.

Seminar February 23–28, 2025 – https://www.dagstuhl.de/25092
2012 ACM Subject Classification Computing methodologies → Search methodologies; Theory of computation → Design and analysis of algorithms
Keywords and phrases estimation-of-distribution algorithms, heuristic search and optimization, machine learning, probabilistic model building
Digital Object Identifier 10.4230/DagRep.15.2.158

1 Executive Summary

John McCall (The Robert Gordon University – Aberdeen, GB)
Josu Ceberio Uribe (University of the Basque Country – Donostia, ES)
Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)
Carsten Witt (Technical University of Denmark – Lyngby, DK)

The Dagstuhl Seminar “Estimation-of-Distribution Algorithms: Theory and Practice” from February 23 to 28, 2025, brought together 25 international experts in the field of estimation-of-distribution algorithms (EDAs). EDAs iteratively evolve a distribution on the search space (called probabilistic model of the search space) in such a way that better solutions accumulate probability mass. Consequently, after running an EDA for sufficient time, one can obtain near-optimal solutions simply by sampling from this probabilistic model. The expertise of the seminar participants covered a broad variety of aspects ranging, in particular, from theoretical topics such as mathematical runtime analyses to an applied topics such as the solution of industrial optimization problems through EDAs.

* Editor / Organizer
† Editorial Assistant / Collector

The main aim of the seminar was to narrow the gap between theory and practice in EDAs by bringing together researchers from both sides and stimulating interaction. We facilitated this interaction through two long introductory talks on theory of EDAs and on typical applications of EDAs, through some regular conference-style talks discussing recent substantial results, short flash talks presenting new ideas and open problems, and breakout sessions for focused discussions. At all times, we left ample time for discussions, and flexibly adjusted the schedule to not cut short interesting discussions.

We believe that the seminar has succeeded its main aims of narrowing the gap between theory and practice in EDAs. This is visible from the high number of talks and the intensive discussions both in plenary and breakout sessions. Moreover, all participants attended essentially all sessions and all stayed for the full five days of the seminar. Given the large number of truly leading experts from academia and industry usually having overloaded schedules, this shows the success of the seminar, and the Dagstuhl Seminar center in general.

The seminar has identified several promising research directions, which can be explored in the reminder of this report. One that deserves particular mention is the relation of model-based genetic algorithms, that use additional mechanisms to understand the search space structure, and multi-variate EDAs, that try to understand the search space via the core working principles of EDAs. Here clearly a more intensive exchange of ideas is very promising, and will surely result from this seminar.

Overall, we feel that this was a very successful seminar, and, also in the name of the participants, we thank Dagstuhl for enabling us to conduct this seminar at Schloss Dagstuhl, which was the ideal place for this type of seminar.

2 Table of Contents

3 Overview of Talks

3.1 Multi-Armed Bandits and Algorithm Configuration and their connections to EDAs (Self-presentation)

Jasmin Brandt (Universität Paderborn, DE)

Main reference Jasmin Brandt, Viktor Bengs, Björn Haddenhorst, Eyke Hüllermeier: "Finding Optimal Arms in Non-stochastic Combinatorial Bandits with Semi-bandit Feedback and Finite Budget", in Proc. of the Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022, 2022.
URL http://papers.nips.cc/paper_files/paper/2022/hash/820e95997d050178323230e316897c38-Abstract-Conference.html
Main reference Jasmin Brandt, Elias Schede, Björn Haddenhorst, Viktor Bengs, Eyke Hüllermeier, Kevin Tierney: "AC-Band: A Combinatorial Bandit-Based Approach to Algorithm Configuration", in Proc. of the Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023, pp. 12355–12363, AAAI Press, 2023.
URL https://doi.org/10.1609/AAAI.V37I10.26456

In this talk, a brief introduction to the Multi-Armed Bandit and Algorithm Configuration problem was given. I presented my recent work on combinatorial bandits and the challenges during the extension of the theoretical analysis to Algorithm Configuration. In addition, the connection of the problems to EDAs was shortly outlined.

3.2 Inversion vectors: an alternative representation for permutations

Josu Ceberio Uribe (University of the Basque Country – Donostia, ES)

Joint work of Josu Ceberio Uribe, Aimar Barrena, Ekhine Irurozki, Jose A. Lozano, Mikel Malagón

Many optimization algorithms use permutations to represent solutions, but despite their apparent simplicity, permutations present significant challenges for optimization methods, particularly for Global Random Search algorithms. The mutual-exclusivity constraint complicates the learning and sampling of probability distributions over the permutation space, often requiring costly numerical computations. An alternative approach involves using bijective functions on $\mathbb{S}_n$ to transform permutation-coded solutions into integer-vector representations known as inversion vectors.

Although inversion vectors have been recognized for centuries, a comprehensive and unified framework for them is lacking. In this paper, we provide formal definitions and a unified notation for all types of inversion vector codifications. We then establish bijective transformations between these codifications, enabling the characterization of their relationships and properties. Finally, we apply this theoretical framework to explain the behavior of UMDA algorithms when different inversion vectors are used in various permutation problems.

3.3 Searching Permutations for Constructing Uniformly Distributed Point Sets

Carola Doerr (Sorbonne University – Paris, FR)

Joint work of François Clément, Carola Doerr, Kathrin Klamroth, Luís Paquete
Main reference François Clément, Carola Doerr, Kathrin Klamroth, Luís Paquete: "Transforming the Challenge of Constructing Low-Discrepancy Point Sets into a Permutation Selection Problem", CoRR, Vol. abs/2407.11533, 2024.
URL https://doi.org/10.48550/ARXIV.2407.11533

Uniformly distributed point sets of low discrepancy are heavily used in experimental design and across a very wide range of applications such as numerical integration, computer graphics, and finance. Recent methods based on Graph Neural Networks [1] and solver-based optimization [2] identified point sets having much lower discrepancy than previously known constructions. We show that further substantial improvements are possible by separating the construction of low-discrepancy point sets into (i) the relative position of the points, and (ii) the optimal placement respecting these relationships. Using tailored permutations, we construct point sets that are of 20% smaller discrepancy on average than those proposed by Rusch et al. In terms of inverse discrepancy, our sets reduce the number of points in dimension 2 needed to obtain a discrepancy of 0.005 from more than 500 points to less than 350. For applications where the sets are used to query time-consuming models, this is a significant reduction.

References

1 T.K. Rusch, N. Kirk, M.M. Bronstein, C. Lemieux, D. Rus, Message-Passing Monte Carlo: Generating low-discrepancy point sets via graph neural networks. *Proc. Natl. Acad. Sci. U.S.A.* 121, e2409913121 (2024). `https://doi.org/10.1073/pnas.2409913121` Free version available at `https://arxiv.org/abs/2405.15059`

2 F. Clément, C. Doerr, K. Klamroth, L. Paquete, Constructing Optimal L_∞ Star Discrepancy Sets. To appear in *Proc. of the AMS*. Preprint available at `https://arxiv.org/abs/2311.17463`

3.4 An Introduction to Search Trajectory Mining

Martin Fyvie (The Robert Gordon University – Aberdeen, GB)

Population-based metaheuristics are commonly used to solve complex optimization problems, but their stochastic nature often obscures decision-making processes. As these algorithms are increasingly deployed in critical domains requiring end-user interaction, the need for explainability has grown. This presentation introduces trajectory mining as a post-runtime analytical framework for enhancing the transparency of optimization methods including Estimation of Distribution Algorithms, Genetic Algorithms, Particle Swarm Optimization, and Differential Evolution.

Our approach is based on the use of Principal Components Analysis to project algorithm search trajectories onto lower-dimensional sub-spaces, capturing population evolution from initialization to convergence. Novel metrics such as Mean Variable Contribution identify decision variables with highest impact on search behaviour and solution quality. These analyses, complemented by statistical testing, extract explanatory features that reveal algorithm behaviour across diverse problem landscapes, allowing search behaviour comparisons.

We show from experiments on benchmark problems and real-world applications how trajectory mining can be used to identify influential variables and reveal key learning steps that help to explain fitness outcomes. These insights can inform the design of new algorithm variants while improving end-user understanding of solution quality and algorithm behaviour, particularly in domains requiring human-algorithm collaboration.

3.5 It's Always the Step-Size

Nikolaus Hansen (INRIA Saclay – Palaiseau, FR)

By the semantics of the parameter K in cGA, we identify analogous places "of K" in the CMA-ES. However, in the CMA-ES, we cannot use the same value for K in all of these places. In particular, in the (m, C) parametrization of the distribution, the C update must have a much larger K value (a smaller learning rate) than the m update, usually n^2/μ vs 1. Otherwise, CMA-ES fails to work. In the CMA-ES, genetic drift is usually only observable when the objective is noisy and we discuss how failures can be observed. Anecdotal evidence also suggests that a failure of ES or CMA-ES is often due to a failure of the step-size adaptation, in particular also when the objective is noisy where we observe the smallest eigenvalue of C to drift to zero exponentially fast.

3.6 Can EDAs be useful for Data augmentation and Imputation?

Jose Ignacio Hidalgo (Complutense University of Madrid, ES)

Joint work of Jose Ignacio Hidalgo, Oscar Garnica, J. Mauel Velasco, Jessica Megane
Main reference José Ignacio Hidalgo, Marta Botella, J. Manuel Velasco, Oscar Garnica, Carlos Cervigón, Remedios Martínez, Aranzazu Aramendi, Esther Maqueda, Juan Lanchares: "Glucose forecasting combining Markov chain based enrichment of data, random grammatical evolution and Bagging", Appl. Soft Comput., Vol. 88, p. 105923, 2020.
URL https://doi.org/10.1016/J.ASOC.2019.105923

In this talk, we present the work of our group on blood glucose prediction for individuals with diabetes, where we utilize Estimation of Distribution Algorithms in two distinct parts of the modeling process. First, we applied a UMDA to learn the weights of an ensemble model for glucose prediction. Second, we also use a UMDA to improve the data augmentation process.

I divided the talk in three parts. In the first part I quickly revised how I used a comatc GA, a basic EDA, for solving the main contribution of my PhD Thesis: Partitioning of digital ciruits for Multi-FPGA implementation. I also mentioned in this part the previous work on EDAs in collaboration qith UPM, Roberto Santana, Pedro Larrañaga, Concha Bielza an Alfredo Cuesta, where we proposed the use of Archimedean Copulas as the model of a Joint Distribution Fuuhction in an EDA.

The second part of the talk was devoted to explain our work on Blood Glucose prediction for peole with diabetes. We have beeb applying Grammatical Evolution approaches for the last 12 years. In those works we use EDAs, in particular an UMDA. First, we applied a UMDA to learn the weights of an ensemble model for glucose prediction. Second, we also use a UMDA to improve the data augmentation process.

Finally, we propose a new problem for EDAs to solve in the context of the EARCO project. EARCO stands for European Alpha-1 Research Collaboration. It is a pan-European network committed to promoting clinical research and education in alpha-1 antitrypsin deficiency (AATD).

3.7 Is Deep Optimization an EDA?

Joshua D. Knowles (SLB Cambridge Research, GB)

Main reference Jamie Caldwell, Joshua D. Knowles, Christoph Thies, Filip Kubacki, Richard A. Watson: "Deep Optimisation: Transitioning the Scale of Evolutionary Search by Inducing and Searching in Deep Representations", SN Comput. Sci., Vol. 3(3), p. 253, 2022.
URL https://doi.org/10.1007/S42979-022-01109-W

Deep Optimization is an algorithm that uses deep association networks to learn epistatic structure in a problem while optimizing it, and that works recursively applying parts of itself to itself. It was introduced in 2018 by Jamie Caldwell et al, [1] and it comes out of and develops ideas from Richard A. Watson's 25 years of research on models of symbiogenetic evolution. It appears to perform remarkably well on some problems (including learning constraints with no explicit constraint handling mechanism), however that's presently a tentative statement because it's not in common use. The basic form of the algorithm is an EDA but it also departs from other EDAs in some ways that I'll explain. I have new results to show based on work I did with MSc students last year; these improve the algorithm significantly on hard multiple knapsack problem instances. Questions: Is Deep Optimization correctly classified as an EDA? Is it better than other EDAs? Why/why not? How should we measure its performance (since it has more and different overhead than other EDAs)? Is there a lighter-weight version of it that would be better for theoretical analysis? What theoretical tools would be needed to analyse Deep Optimization properly?

References

1 Caldwell, J.R., Watson, R.A., Thies, C. and Knowles, J.D., 2018. Deep optimisation: Solving combinatorial optimisation problems using deep neural networks. arXiv preprint arXiv:1811.00784.
2 Caldwell, J., Knowles, J., Thies, C., Kubacki, F. and Watson, R., 2022. Deep optimisation: Transitioning the scale of evolutionary search by inducing and searching in deep representations. SN Computer Science, 3(3), p.253.

3.8 Estimation-of-Distribution Algorithms for Multi-Valued Decision Variables

Martin S. Krejca (Ecole Polytechnique – Palaiseau, FR)

Joint work of Firas Ben Jedidia, Benjamin Doerr, Martin S. Krejca
Main reference Firas Ben Jedidia, Benjamin Doerr, Martin S. Krejca: "Estimation-of-distribution algorithms for multi-valued decision variables", Theor. Comput. Sci., Vol. 1003, p. 114622, 2024.
URL https://doi.org/10.1016/J.TCS.2024.114622

The majority of research on estimation-of-distribution algorithms (EDAs) concentrates on pseudo-Boolean optimization and permutation problems, leaving the domain of EDAs for problems in which the decision variables can take more than two values, but which are not permutation problems, mostly unexplored. To render this domain more accessible, we propose a natural way to extend the known univariate EDAs to this setting. Different from a naive reduction to the binary case, our approach avoids additional constraints. Since understanding genetic drift is crucial for an optimal parameter choice, we extend the known quantitative analysis of genetic drift to EDAs for multi-valued, categorical variables. Roughly speaking, when the variables take r different values, the time for genetic drift to become significant is r times shorter than in the binary case. Consequently, the update strength of the probabilistic model has to be chosen r times lower now. To investigate how desired model updates take place in this framework, we undertake a mathematical runtime analysis on the r-valued LeadingOnes problem. We prove that with the right parameters, the multi-valued UMDA solves this problem efficiently in $O(r \ln(r)^2 n^2 \ln(n))$ function evaluations. Overall, our work shows that our good understanding of binary EDAs naturally extends to the multi-valued setting, and it gives advice on how to set the main parameters of multi-values EDAs.

3.9 Introduction to Theoretical Analyses of EDAs

Timo Kötzing (Hasso-Plattner-Institut, Universität Potsdam, DE)

The Theory of EDAs, conducted as a run time analysis of specific EDAs applied to test functions, has a history of close to 20 years. In this talk, I review some of the core results provided by the theory of EDAs. I give insights into main paradigms and the models used by theoreticians.

Discussing results, I present different settings pertaining to (a) hill climbing; (b) noise, rugged landscapes and local optima; (c) behavior in the absence of a signal.

3.10 Faster Optimization Through Genetic Drift

Johannes Lengler (ETH Zürich, CH)

Joint work of Cella Florescu, Marc Kaufmann, Johannes Lengler, Ulysse Schaller, Dirk Sudholt, Carsten Witt
Main reference Cella Florescu, Marc Kaufmann, Johannes Lengler, Ulysse Schaller: "Faster Optimization Through Genetic Drift", in Proc. of the Parallel Problem Solving from Nature – PPSN XVIII – 18th International Conference, PPSN 2024, Hagenberg, Austria, September 14-18, 2024, Proceedings, Part III, Lecture Notes in Computer Science, Vol. 15150, pp. 70–85, Springer, 2024.
URL https://doi.org/10.1007/978-3-031-70071-2_5

I gave some new perspective on genetic drift. While genetic drift is often seen as something to be avoided, as it produces mistakes, I discuss a simple problem, dynamic BinVal, for which it is faster to accept those mistake and use aggressive updates that correct them quickly. This talk is part of a general discussion at Dagstuhl on when one should avoid or accept genetic drift.

References

1 Johannes Lengler, Dirk Sudholt, and Carsten Witt. "The complex parameter landscape of the compact genetic algorithm." Algorithmica 83 (2021): 1096-1137.

3.11 Introduction to EDAs

John McCall (The Robert Gordon University – Aberdeen, GB)

In this talk I provide an introduction to Estimation of Distribution Algorithms (EDA). I begin with the general concept of an EDA: that one models the distribution of fitness over a search space and by sampling this distribution one hopes to generate high fitness solutions with a high probability.

The talk covers a range of key EDA techniques including: linkage learning; model parameterisation, model sampling; and adaptation to different representation spaces. I also present some EDA applications on test functions, theoretical benchmarks and to real world applications in biocontrol and marine scheduling.

In summary, EDAs are a mature and successful optimisation metaheuristic based on mathematically-principled statistical modelling that have been applied across all major problem representations and in the real world.

3.12 Estimation of Distribution Algorithms using generative Machine Learning models

Franz Rothlauf (Universität Mainz, DE)

The talk discussed the development of EDAs using generative ML models over the last 15 years.

3.13 From Fitness Landscapes to Problem Factorizations: A Machine Learning Perspective

Roberto Santana (University of the Basque Country – Donostia, ES)

Joint work of Roberto Santana, Arnaud Liefooghe, Bilel Derbel
Main reference Roberto Santana, Arnaud Liefooghe, Bilel Derbel: "Boomerang-shaped neural embeddings for NK landscapes", in Proc. of the GECCO '22: Genetic and Evolutionary Computation Conference, Boston, Massachusetts, USA, July 9 – 13, 2022, pp. 858–866, ACM, 2022.
URL https://doi.org/10.1145/3512290.3528856

This talk discusses the use of machine learning models – particularly neural networks – to automate various components of evolutionary search by leveraging information from the problem's fitness landscape. A general framework is presented in which the search model comprises three elements: (1) data from the fitness landscape; (2) a neural network; and (3) a machine learning task. As a specific instance of this framework, the talk presents the problem of learning neural embeddings for NK landscapes [1]. In this scenario, the goal is to learn an embedding-based representation of an NK landscape instance, where each binary solution of the NK problem is mapped to a continuous vector. To construct this embedding, each NK model solution is treated as a word in a vocabulary. Sentences are sequences of NK solutions related by predefined semantics (e.g., walks in the fitness landscape), and a corpus is a large collection of such sentences.

The talk also proposes alternative scenarios in which fitness landscape data correspond to local optima networks (LONs) [2], represented as annotated graphs. Graph neural networks are employed to process these LONs, and the machine learning task involves algorithm parameter generation by producing a factorization of the optimization problem – one that could be exploited by a factorized distribution algorithm (FDA). The work aligns with approaches that integrate ML techniques with classical exploratory landscape analysis for automated algorithm selection [3], as well as with applications of graph neural networks in solving combinatorial optimization problems [4].

Finally, the talk addresses the question of theory choice [5] for Estimation of Distribution Algorithms, advocating for the use of factorizations as a fundamental component of EDA theory. This is justified by their potential to provide simplicity, broad applicability, fruitfulness, and accuracy – particularly due to their compatibility with widely accepted theories in the ML field.

References

1 R. Santana, A. Liefooghe, and B. Derbel, "Boomerang-shaped neural embeddings for nk landscapes," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2022, pp. 858–866.
2 G. Ochoa, M. Tomassini, S. Vérel, and C. Darabos, "A study of nk landscapes' basins and local optima networks," in *Proceedings of the 10th annual conference on Genetic and evolutionary computation*, 2008, pp. 555–562.
3 M. Seiler, U. Škvorc, C. Doerr, and H. Trautmann, "Synergies of deep and classical exploratory landscape features for automated algorithm selection," in *International Conference on Learning and Intelligent Optimization.* Springer, 2024, pp. 361–376.
4 I. Echeverria, M. Murua, and R. Santana, "Multi-assignment scheduler: A new behavioral cloning method for the job-shop scheduling problem," in *International Conference on Learning and Intelligent Optimization.* Springer, 2024, pp. 138–152.
5 T. S. Kuhn, *The Essential Tension.* University of Chicago Press, 1977, ch. Objetivity, value judgment and theory choice, pp. 320–339.

3.14 The use of the Doubly Stochastic Matrix models for the Quadratic Assignment Problem

Valentino Santucci (University for Foreigners of Perugia, IT)

Joint work of Valentino Santucci, Josu Ceberio
Main reference Valentino Santucci, Josu Ceberio: "On the Use of the Doubly Stochastic Matrix Models for the Quadratic Assignment Problem", Evolutionary Computation, pp. 1–33, 2025.
URL https://doi.org/10.1162/evco_a_00369

Permutation problems have captured the attention of the combinatorial optimization community for decades due to the challenge they pose. Although their solutions are naturally encoded as permutations, in each problem, the information to be used to optimize them can vary substantially. Therefore, we discussed two types of permutation problems: ordering problems and assignment problems. The Quadratic Assignment Problem (QAP) is perhaps the most prominent example of permutation problems of assignment type. In the talk we discussed how it is possible to use Doubly Stochastic Matrices (DSMs) to address QAP in the context of Estimation of Distribution Algorithms. To that end, we discussed efficient learning and sampling schemes that enable an effective iterative update of the probability model. We analyzed the effectiveness of the discussed model both on QAP and on the the Linear Ordering Problem (LOP), a well-known example of permutation problem of ordering type. The talk concluded with a description of the potential uses of DSMs for other optimization paradigms, such as genetic algorithms or model-based gradient search.

3.15 Linkage-Learning Evolutionary Algorithms: Recent Theoretical Advances

Marcus Schmidbauer (Universität Passau, DE)

Joint work of Markus Schmidbauer, Dirk Sudholt

Linkage Learning (LL) aims at detecting dependencies between problem variables. We gave a short introduction into LL and, in particular, contrasted different linkage-learning techniques. Statistical Linkage Learning (SLL) uses frequencies of gene value combinations, whereas Empirical Linkage Learning (ELL) is based on investigating how perturbation influences the fitness of a solution. We proved the correctness of ELL on the Hierarchical If-and-only-if (H-IFF) problem. Further, we stated bounds on the expected runtime of the ELL-based Parameter-less Population Pyramid (P3) on H-IFF.

3.16 Estimation of Distribution Algorithms and the Ising Model

Jonathan L. Shapiro (University of Manchester, GB)

This talk discussed Estimation of Distribution Algorithms and the Ising Model.

3.17 Vine Copula-Based Estimation of Distribution Algorithms for Multivariate Continuous Optimization

Marta Rosa Soto Ortiz (CaixaBank Tech – Madrid, ES)

Estimation of Distribution Algorithms (EDAs) are a class of generative models within evolutionary computation that build probabilistic models from selected solutions in order to generate new candidate solutions. In the continuous multivariate setting, traditional EDAs often assume Gaussian distributions and linear dependencies between variables. However, these assumptions are restrictive and rarely hold in complex, real-world scenarios.

To address these limitations, we adopt vine copula models, which offer a flexible framework by decoupling the modeling of marginal distributions from the dependency structure. Vine copulas capture dependencies between variables – regardless of their individual distributions – by learning them directly from the data, allowing for a broad range of pairwise relationships, including asymmetric, nonlinear, and non-monotonic interactions.

Moreover, vine copulas decompose the joint dependency structure into a product of conditional bivariate copulas, organized hierarchically in a sequence of trees known as R-vines. This architecture allows different copula families to be assigned to different variable pairs within the same decomposition, making it highly adaptable to the complex dependency patterns found in real-world data.

We illustrate the effectiveness of this approach through real-world examples, where vine copulas successfully capture diverse dependency patterns – such as tail dependence and complex variable dependencies – that Gaussian-based models fail to represent. In addition, we introduce copulaedas, an R package available on CRAN, which implements a complete vine copula-based EDA framework, including model learning and sampling procedures.

Our findings show that vine copulas significantly enhance the modeling capabilities of EDAs, enabling more expressive and adaptive search distributions for continuous optimization problems.

3.18 Optimal Mixing in Model-based Evolutionary Algorithms

Dirk Thierens (Utrecht University, NL)

This talk discussed Optimal Mixing in Model-based Evolutionary Algorithms.

3.19 Estimated Distributions and Warm Starts

Vanessa Volz (CWI – Amsterdam, NL)

Estimation-of-Distribution Algorithms (EDA) learn the underlying structure of a given black-box problem. But can the obtained model be reused for optimising similar problems? Below, I summarise a discussion on the topic.

In the field of evolutionary computation (EC), problems are commonly assumed to be a black box. However, in most real-world settings, solving a problem with sophisticated methods is only considered profitable if it is *recurring*, i.e. if different instances of the same problem are encountered repeatedly. For example, logistics companies need to solve comparable routing problems every day. Medical professionals need to create effective treatment plans for similar conditions. Hyperparameters of algorithms need to be adjusted for each new task. While the underlying problem structures remain unknown, black-box optimisation approaches could be iteratively tailored to the regularly encountered problem over time (across instances). Essentially: *Learning to optimise*.

Some work on warm-starting EDAs exists, for example for CMA-ES ([1] for Hyperparameter Optimisation and [2] for contextual optimisation). A structural transfer framework using EDAs has also been proposed in [3].

For some problems, such as ensemble selection of machine learning models for different data folds, the optima may be assumed to be in similar locations. In these cases, when running an EDA, warm-starting from the previously estimated distribution (with higher exploration settings) might be sufficient for improving algorithm performance. In other cases, additional domain knowledge would be required to determine how the obtained information can be transferred. However, in most settings, what kind of similarity exists between instances would need to be established first. How this can be done remains an open question.

References

1 M. Nomura, S. Watanabe, Y. Akimoto, Y. Ozaki, and M. Onishi, *Warm Starting CMA-ES for Hyperparameter Optimization*. In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 10, Art. no. 10, May 2021, doi: 10.1609/aaai.v35i10.17109.

2 Y. Sekino, K. Uchida, and S. Shirakawa, *Warm Starting of CMA-ES for Contextual Optimization Problems*. In Parallel Problem Solving from Nature – PPSN XVIII, Cham: Springer Nature Switzerland, 2024, pp. 205–220. doi: 10.1007/978-3-031-70068-2_13.

3 R. Santana, A. Mendiburu, and J. A. Lozano, *Structural transfer using EDAs: An application to multi-marker tagging SNP selection*. In 2012 IEEE Congress on Evolutionary Computation, Jun. 2012, doi: 10.1109/CEC.2012.6252963.

3.20 Runtime Analysis of the Multi-valued cGA

Carsten Witt (Technical University of Denmark – Lyngby, DK)

Joint work of Sumit Adak, Carsten Witt
Main reference Sumit Adak, Carsten Witt: "Improved Runtime Analysis of a Multi-Valued Compact Genetic Algorithm on Two Generalized OneMax Problems", CoRR, Vol. abs/2503.21439, 2025.
URL https://doi.org/10.48550/ARXIV.2503.21439

In this talk, we discuss recent advances in the runtime analysis of the multi-valued cGA, defined in the framework by Ben Jedidia et al. [5]. We consider an r-valued LeadingOnes function and prove a runtime bound of $O(n^2r^2(\log^3 n+\log^2 r))$ that holds with high probability. For an r-valued OneMax function counting the number of positions of value $r-1$, we prove a bound of $O(nr(\log n + \log^2 r))$, which improves the previously best bound by Hamano et al. [4] by a logarithmic factor and is tight for $r = 2$. Finally, we discuss a second generalization of OneMax to the r-valued domain, which sums up the integer values of all positions and hence has a maximum value of $n(r-1)$. For this function, the multi-valued cGA is proved to find the optimum in time $O(nr^3(\log^2 n + \log r))$ with high probability. Our proofs are based on analyses of stochastic drift of the frequencies for value $r-1$ and bounds on their genetic drift. Finally, we discuss why the bound for the second OneMax-function is probably not tight.

References

1 S. Adak and C. Witt, *Runtime Analysis of a Multi-valued Compact Genetic Algorithm on Generalized OneMax*, in Proc. of PPSN 2024, Springer, pp. 53-69.
2 S. Adak and C. Witt, *A Runtime Analysis of the Multi-Valued Compact Genetic Algorithm on Generalized LeadingOnes*, in Proc. of EvoCOP 2025, Springer, to appear.
3 S. Adak and C. Witt, *Improved Runtime Analysis of a Multi-Valued Compact Genetic Algorithm on Two Generalized OneMax Problems*, in Proc. of GECCO 2025, ACM Press, to appear
4 R. Hamano, K. Uchida, S. Shirakawa, D. Morinaga, and Y. Akimoto, *"Tail Bounds on the Runtime of Categorical Compact Genetic Algorithm*, Evolutionary Computation, 2025, `https://doi.org/10.1162/evco_a_00361`, in press
5 F. Ben Jedidia, B. Doerr, and M. S. Krejca, *Estimation-of-Distribution Algorithms for Multi-Valued Decision Variables*, Theoretical Computer Science, vol. 1003, p. 114622, 2024.

4 Working groups

4.1 Breakout Session: Learning Distributions with Deep Learning

Josu Ceberio Uribe (University of the Basque Country – Donostia, ES)

The breakout session explored applications of deep learning for learning distributions, with a focus on alternatives to traditional Maximum Likelihood Estimation (MLE) approaches. Starting point of the session was the following:

- Traditional MLE methods often require intensive computation,
- Model capacity must appropriately capture variable interactions,
- Deep learning offers computational efficiency while maintaining modeling power.

There was no common position, but rather, each participant presented a different vision of what the title of the session might mean. As a result, participants suggested various approaches:

1. Generating variation rather than modeling entire populations,
2. Moving beyond optimization to distribution estimation,
3. Finding optimal search directions,
4. Leveraging deep learning models (both custom-trained and existing LLMs) to guide search point selection,
5. Developing higher-level learning models adaptable to various problems,
6. Using fitness approximation for ranking or selection,
7. Integrating selection directly into learning processes using fitness information.

Finally, the conversation expanded to more general ideas:

- Learning movement patterns or genetic operators (crossover, mutation),
- Developing black-box iterative models that learn movement, selection, and solution representation,

- Projection techniques into embedding spaces to reduce interactions while preserving information,
- Relationships with Bayesian Optimization,
- Solution quality prediction and generation (potential GAN applications).

4.2 Breakout Session: Exploration versus Exploitation

Duc-Cuong Dang (Universität Passau, DE), Jonathan L. Shapiro (University of Manchester, GB)

Participants. Jonathan L. Shapiro, Franz Rothlauf, Carsten Witt, Vanessa Volz, Jasmin Brandt, Timo Kötzing, Martin S. Krejca, Aishwarya Radhakrishnan, Andrew M. Sutton, Marcus Schmidbauer, Duc-Cuong Dang

To get everyone on the same page, Jonathan Shapiro highlighted the importance of exploitation and exploration as in the case of Reinforcement Learning (RL): at any time one always have to decide between using the current information to exploit immediate rewards, versus exploring the alternatives that may lead to better rewards later.

The first question we discussed is whether it is a good strategy to explore first and exploit later. This is the case with traditional Evolutionary Algorithms (EAs), the algorithms often converge in the end. Some participants noted that this is not the case for EDAs like BoA. Marcus Schmidbauer added that in the case of GOMEA, a multivariate model-building algorithm, it is a combination of extreme exploitation with extreme exploration. Timo Kötzing added a comment that there is a fundamental difference between RL and EA: in EAs, the rewards are not immediate. The notion of neighbourhood is also important: EAs seem to prefer having nearby good solutions over having random good solutions that are hard to recombine. Jasmin Brandt commented that we may need a proper definition of exploitation and exploration. According to Franz Rothlauf, exploitation means finding good solutions nearby the ones already explored, while doing anything else is exploration.

We then discussed different parameters that can influence the search towards either exploitation or exploration. In the case of EAs, it appears that that the mutation rate is often settled to one on average per mutation. However, one should be careful about which EAs are considered. In the case non-elitist EAs, the balance between mutation and selection does not necessarily lead to mutation rate of $1/n$ but it depends on which selection is used or how it is set. For example, in tournament selection of size k the balance is found exactly at mutation rate $\ln(k)/n$. For EDAs, the following parameters can influence the search: the inverse of the population size $1/K$ in cGA; for UMDA, this is the ratio μ/λ when the (μ,λ)-truncation selection is used to decide which solutions to keep in order to model the next distribution; the learning rate ρ in PBIL and related algorithms like Ant Colony Optimisation. From the theory side, how to control these parameters optimally for simple problems remains an open question as the existing lower bound techniques and results are scarce.

A common set of parameters that is important for all the above-mentioned algorithms to avoid the possibility of premature convergence is the borders that limit the frequencies of bits to be strictly below 1 and strictly above 0. In practical applications, the use of these borders is essential as one may absolutely wants to avoid degenerated distributions, an example is the application Franz Rothlauf has in his talk. Related to this, Jonathan Shapiro asked, in practices how to avoid identify that the learned model is general enough and not over-fitted. This is a hard question, and often the case this comes back to a good engineering practice, or a familiarity with the problems and algorithms. Martin S. Krejca put forwards the following question: Is there any way to adjust the borders, somehow reflecting the amount of confidence on doing the right thing at any given time? Jasmin Brandt added that in the case of multi-arm bandit problems, the confidence interval is a well-defined concept. A related question we discussed is how to detect that the entropy of the model is too low early on, or are the techniques developed for stagnate detection in EAs useful for EDAs?

In summary, we all agree that the balancing exploitation and exploration is relevant to all areas of Evolutionary Computation, including EDAs. We know how to deal with them for practical problems, however formulating them as statements with rigorous proofs remains challenging. The good news is that even in a lack of recommendations, there is a technique from theory that allow selecting the values for parameters efficiently based on power-law distributions.

4.3 Breakout Session: Univariate versus Multivariate 2

Duc-Cuong Dang (Universität Passau, DE), Jonathan L. Shapiro (University of Manchester, GB)

Participants. Jonathan L. Shapiro, John McCall, Joshua D. Knowles, Carola Doerr, Vanessa Volz, Nikolaus Hansen, Josu Ceberio Uribe, Valentino Santucci, Alistair Benford, Andrew M. Sutton, Martin S. Krejca, Roberto Santana, Duc-Cuong Dang

The breakout session began with a discussion of the trade-off between having an accurate model, at the expense of intensive computation, as in the case of building graphical models in EDAs, versus having less accurate models that are faster to learn, but at the risk of misleading or slowing down the search for optimal solutions.

We agree that in practical runs of EDAs, there may not be a clear separation between the phase of learning an accurate model and that of using that model to find optimal solutions as the model is often learned along the way. More importantly, as pointed out by Nikolaus Hansen, for solving an optimisation problem, finding an accurate model can be misleading: we only want to find the best solutions. In fact, we want models that can help the search in finding better solutions, or in finding improvements to the existing solutions, more efficiently. In other words, a good model should help guide the search towards the optimal solutions, and need not be an accurate description of those solutions.

We then discussed whether there are practical problems where there are clear interactions between variables, but the univariate model is still useful, as this is often argued by theory research (most existing theory papers only focus on univariate models). John McCall gave an example of such settings for a scheduling problem in chemotherapy. In this application, thousands of interactions between variables are detected between the variables, but univariate algorithms like PBIL/DEUM still outperform hBOA. The current hypothesis for the inefficiency of hBOA, which is stuck with medium quality solutions, is that this is due to the dominance of different terms in the objectives. It could be very interesting if there is a mathematical description that can capture this property.

As the discussion progressed, we asked the question whether the computational budget or parameter settings, like the sample size, can have some influence on the success of univariate models. This is because the univariate model is cheaper to learn, compared to models with interactions. From the practical side, practitioners have noticed that on some problems, the population or the sample size needs to be set above a certain threshold for the search to be successful. We then discussed the settings in which the computational budget is less of an issue, whether multivariate models are always the better choice. Nikolaus Hansen mentioned a counter-example of the sector-sphere function where univariate performs best. In addition, trying to adapt the step sizes on this function may worsen the progress of optimisation. Similar effects can be observed for the case of LongPath and SurfSamp in discrete optimisation, and this is related to the general issue of balancing exploitation and exploration.

We then discussed whether multivariate models might limit the diversity of solutions. In other words, in multivariate EDAs, how does the entropy of the model evolve as the search progresses? Nikolaus Hansen said that it tends to increase as we learn more about the models. He added that a good indicator of a good model is that the entropy of sampled solutions within a fitness level should be high. Carola Doerr drew the parallel with black-box complexity: if we know more about the problem at hand, we want to have higher entropy. For example, if we know that we are solving a OneMax-type problem, the best algorithm is to sample $O(n/\log n)$ solutions uniformly at random, and then infer a consistent target string based on the acquired fitness of these random solutions.

The following questions were also discussed during the breakout session.

1. Is it always a good practice to first try univariate models first and then try the multivariate ones? There is no consensus among the participants on this question.
2. What is the current state of theory research on multivariate models? As far as we know, the only theory research on multivariate models is about linkage-learning, e.g. the papers by Dirk Sudholt and co-authors.
3. Should we model improvement moves (or directions) instead of that of solutions? After our discussion, we incline towards modelling the improvement moves.
4. Is there any benefit for building a model from scratch instead of incremental learning, and vice-versa? This is open.
5. Are the weights of detected interactions important to learn, and how to use them? This is open.
6. Is there a way to learn the model implicitly, or somehow avoid the burden of building an explicitly accurate model? This is open, but there are related ideas about projecting the search into other spaces where there may be less deception.

4.4 Breakout Session: Genetic Drift 1

Martin S. Krejca (Ecole Polytechnique – Palaiseau, FR)

Participants. Alistair Benford, Jasmin Brandt, Josu Ceberio Uribe, Duc-Cuong Dang, Benjamin Doerr, Carola Doerr, Martin Fyvie, Nikolaus Hansen, Joshua Knowles, Timo Kötzing, Martin S. Krejca

This breakout session discussed the phenomenon of genetic drift in estimation-of-distribution algorithms (EDAs). The discussion meandered a bit while it took place. The following summary rearranges it logically, grouping it in terms of different points that were discussed.

We started the session by Benjamin's explanations of the mathematical results for genetic drift in univariate EDAs on binary problems, which mostly concerns the cGA and the PBIL. In this setting, genetic drift refers to the random movement of frequencies in the probabilistic model without any true signal from the objective function. In the case of constant objective functions, we have strong concentration bounds that tell us how long it takes until a frequency randomly moves by at least ε. It was also mentioned that the resulting bounds can be used for a smart restart scheme, which recommends to restart an EDA whenever it is likely that a frequency could have moved too far due to genetic drift. When restarting, the step size of an update should be halved (which usually means that the population size should be doubled).

This explanation raised the question how realistic this analysis is, since real objective functions are not constant. It was clarified that genetic drift is usually superimposed by the true signal of an objective function, but that it is hard to analyze it in its generality, which is why the mathematical analysis assumes a constant objective function, which represents the full effect of pure genetic drift.

A natural question that followed was whether this phenomenon also shows up in practice and whether it is relevant. It was agreed upon that genetic drift does show up in practice, but it was less clear whether it is relevant. The consensus was ultimately that is is indeed relevant but that practitioners may not care about it due to other, more pressing issues, such as budgetary constraints or the wishes of the customer. A typical approach is to simply rely on best practices, which traditionally aim at reducing the impact of genetic drift (albeit without mathematical guarantees).

Afterward, the discussion moved to other types of search spaces. It was mentioned that genetic drift is also present for multi-valued discrete search spaces, where it is even easier to have a greater (detrimental) effect, due to generally smaller starting values of the frequencies. Afterward, the question came up to what extent genetic drift is visible in continuous search spaces. The conclusion was that it also shows up there and also poses a problem.

With genetic drift being a general problem, other questions were whether it is possible to detect genetic drift somehow and how it behaves in the presence of noise. Regarding the detection, it was mentioned that the sig-cGA operates on only updating a frequency when a statistical significance in the objective function is detected. This approach allows to directly measure (with a certain confidence) that genetic drift did *not* occur. This approach should not be largely affected by noise. Otherwise, more traditional EDAs are known to cope well with noise, but only if the step size is chosen sufficiently small (suggesting again a smart restart scheme in the general case).

These points were followed up by questions how genetic drift relates to similar phenomena in other settings, such as multi-armed bandits and reinforcement learning. For multi-armed bandits, it was not so clear how this directly translates to the setting of EDAs, since there is no real exploitation in the binary domain, as the two possible choices for a variable are either 0 or 1. However, assigning a confidence to each frequency is something that aligns well with multi-armed bandits and seems like a good strategy. As for reinforcement learning, the question was whether the approach of adding uniform samples periodically with a certain probability could prove promising to counteract genetic drift. The argument was brought forth that uniform samples are not a good choice. However, in continuous problems, it is not uncommon to add noise on purpose to the samples and to then assess whether the selection is affected by this choice. If it is, then the learning rate is increased. Something similar could potentially prove useful for EDAs as well.

In the same vein as the connection to reinforcement learning, the question came up whether it is smart to introduce an artificial bias into the update that explicitly counteracts genetic drift. This has been done before in theoretical studies to great effect, but the paper on the sig-cGA shows that this can also easily backfire. A static approach is not useful, as the bias can be stronger than the signal of the objective function, which renders optimization unsuccessful. However, a dynamically adjusting approach could maybe work.

In conclusion, genetic drift is a real and important problem that is hard to detect. Being able to detect it could tremendously improve the situation. The sig-cGA as well as the concept of adding noise on purpose and adjusting the step size based on whether this affects selection can be promising ways to reduce the impact of genetic drift. However, these details need to be worked out in greater detail analytically.

4.5 Breakout Session: Univariate versus Multivariate 1

Timo Kötzing (Hasso-Plattner-Institut, Universität Potsdam, DE)

In this breakout session we discussed univariate versus multivariate EDAs.

We started by discussing insights dating back to 80s / 90s. In particular, we discussed a concatenated (summed) Trap_k function, and the difference to LeadingOnes-concatenated Trap_k. Univariate EDAs or uniform crossover can optimize this function if the gap in function value between local optima is large enough. It might also work if parameter scale with k sufficiently, due to averaging over many traps. If the gap in fitness is small, the function is "fully deceptive", univariate EDAs get stuck, same with crossover. As an aside: univariate EDAs might be "more stuck" than EAs. Mutlivariate EDA can learn blocks; can swap all blocks. This is highly related to the work of Mühlenbein, Ochoa and others.

We then discussed how to learn the linkage. How can one do it efficiently? The direct algorithm scales quadratic in the number of variables, which can be too much. Many problems with low computation time use UMDA instead. Some modern algorithms can do better (eg GOMEA).

4.6 Breakout Session: Genetic Drift 2

Johannes Lengler (ETH Zürich, CH) and Carsten Witt (Technical University of Denmark – Lyngby, DK)

This 2nd breakout session on genetic drift considered its definition, theoretical implications, and practical considerations in various types of EDAs. The participants discussed how genetic drift affects algorithm performance, its distinction to stochastic drift, and potential strategies for mitigating its impact.

The following **key points** were debated in the session.

Definition and Characteristics of Genetic Drift

Genetic drift was described as the loss of alleles in a population due to stochastic effects. In an infinite population model, genetic drift cannot lead to a loss of alleles. To avoid too strong genetic drift, the population size of an EDA usually has to scale linearly in problem size on benchmark problems like BV (Binary Value) and LO (LeadingOnes). This was already observed in early literature.

Distinguishing Genetic Drift from Stochastic Drift

Empirical evidence suggests sharp thresholds in population size that determine genetic drift behavior. The role of minimum population size in modeling multivariate problems was debated. The impact of genetic drift on neutral versus non-neutral bits was questioned.

Genetic Drift in Specific Problem Settings

For constrained problems (e.g., knapsack), the effect of different selection operators on drift as discussed. Ising models (related to so-called checkerboard problems) were proposed as potential test cases. Whether genetic drift occurs on fitness plateaus was questioned. The Dynamic BinVal problem was suggested as another theoretical test bed.

Counteracting Genetic Drift

In theoretical studies, borders on marginal frequencies are used to avoid fixation at extreme values. Multi-factor EDAs use Bayesian priors to prevent premature fixation. Tournament selection in UMDA and its potential impact on genetic drift was explored. Hybrid approaches combining EDAs with local search were proposed to mitigate drift. Finally, theoretical studies have identified scenarios (dynamic BinVal) where genetic drift can speed up optimization.

Algorithmic and Theoretical Considerations

Studying PBIL instead of UMDA was suggested as potentially more insightful. Again, the role of population size in determining genetic drift was highlighted. The phenomenon of domino convergence was discussed in relation to problem landscapes like BinVal and OneMax.

Broader Implications and Open Questions

The applicability of genetic drift to machine learning was debated. Whether overparameterized neural networks exhibit genetic drift was questioned. Symmetry breaking in EDAs and theoretical analysis of PBIL for feature selection were suggested as future research directions.

4.7 Breakout Session: Explainability of EDAs

John McCall (The Robert Gordon University – Aberdeen, GB) and Josu Ceberio Uribe (University of the Basque Country – Donostia, ES)

In this session, we approached the "explainability" of EDAs from the point of view of expert practitioners. That is to say, what can be derived empirically from the EDA process that can potentially be useful in gaining an understanding of an EDA? The key starting point for the discussion was the observation that runs of an EDA using a Probabilistic Graphical Model (PGM) will generate sequences of PGMs. These can be considered as trajectories of the EDA through model space. What might be learned from these trajectories?

The trajectories are a function of EDA parameters (population size, selection pressure) and problem characteristics (the distribution of fitness over solutions). Generally, insights can be gained from varying algorithm or problem characteristics and deriving statistical relations.

Problems can also be decomposed using certain functional bases on solution spaces (Walsh, continuous Fourier, discrete Fourier). There is rich potential for exploiting this when the PGM explicitly distributes fitness over these basis components, also known as "problem structure."

In black-box scenarios, where the problem structure is unknown and so there is no decomposition available, does it make sense to capture first-order structure, then second-order, and so on to higher orders? Structure detection algorithms are widely used in EDAs and so there is a potential value in mining PGM trajectories for problem structure. We wonder if there is an algorithm that tries to guess the most interesting representation to optimize a problem.

How can a decomposition of the above help when designing an algorithm? Can structural features of problem classes suggest EDAs specialized to learn how fitness is distributed across those structures? Might the behavior of such EDAs on such problems be amenable to theoretical analysis?

EDA experts in the group observed that by examining PGM trajectories you can sometimes manually identify solutions that are better than those sampled from the model by the EDA, even after parameter optimization. Can we learn more efficiently from those trajectories?

In the context of real-world problems, often similar problems need to be frequently solved. We discussed how to gain knowledge about the problem family by running EDAs on the problem and extracting recurrent model structures. Related to this, transfer learning works on EDAs, and incremental-BOA were discussed. There were cases where models learnt in certain iterations were re-used in later iterations (i+1, i+2, i+3,...). We discussed aggregating trajectories of EDA executions, and then using SVD-type techniques to remove noise, and hence gain insights about recurrent structures. We discussed whether ensembles of algorithms (the same algorithm with different parameters run on the same problem) might help in problem understanding. This was related to Island models. In some real-world problems such as scheduling, such analysis can be used to identify resources that need to be carefully allocated early in the solution process, while others may be fixed later and can be flexibly allocated. This also gives a lot of information about the process. We discussed the general concept of "backbones" of important variables that, once optimally set, reduce the complexity of optimizing the remaining variables.

We discussed the relation between EDA theory and EDA explainability, with a special focus on the phase transition between divergence and convergence of algorithms and the relation to population size. A key theme arising many times in the seminar was the focus of EDA theory on genetic drift whereas researchers of large-scale and applied EDAs never consider but see premature convergence as a major problem. The process of adjusting population size appears to be analogous to the role of K in theory. Is there potential for combining theoretical and empirical analysis to understand this phase transition?

In the context of UMDA (or PBIL), we considered statistical analysis of univariate probabilities as a source of information about the problem. Does the explainability approach discussed in this group offer a bridge between the two EDA communities to build a framework that is useful to both theory and practice? Such a framework would consider a set of problems with known and controllable structures and explore generalized concepts of K theoretically and empirically.

4.8 Breakout Session: Linkage Learning

John McCall (The Robert Gordon University – Aberdeen, GB) and Duc-Cuong Dang (Universität Passau, DE)

Participants. John McCall, Marcus Schmidbauer, Benjamin Doerr, Timo Kötzing, Martin Fyvie, Aishwarya Radhakrishnan, Andrew M. Sutton, Nikolaus Hansen, Carsten Witt, Duc-Cuong Dang.

The group considered the question "How much precision is needed in linkage learning?" More fully, for an EDA that learns linkage, either in the form of a PGM (Probabilistic Graphical Model) or more broadly as families of subsets of variables, to optimise a problem, is it necessary to incorporate in the EDA model all linkage that is detected.

The discussion focussed around a published paper comparing algorithm performance on a real world chemotherapy problem. The chemotherapy problem is based on binary strings encoding dosages of multiple drugs over a time period. The problem contains thousands of variable linkages based on tumour-size, dosage constraints and toxic side-effects on the patient. The paper compares a leading multivariate-PGM EDA, hBOA, against a GA and three univariate EDAs (UMDA, PBIL and a univariate version of DEUM).

The counter-intuitive result is that the univariate algorithms, led by UMDA, all converge to the (best-known) optimum much faster than hBOA, with far fewer fitness evaluations and massively less clock time. This is despite the fact that the univariate EDAs learn no interactions. The GA is outperformed on this problem by all of the EDAs.

There was discussion on the linkage detection method used – if there is no thresholding then spurious linkages can be detected that have no significance for the problem. hBOA uses internal linkage parameters that do use thresholding, and note that the inventor of hBOA (Martin Pelikan) collaborated on the paper. Moreover, a separate detection algorithm (LDA) was used to verify the linkages. The paper concludes that, in this example, multivariate linkage is inessential.

A contrasting example in the case of Concatenated Traps was discussed. Here it can be shown theoretically that P3 (Parameter-less Population Pyramid), a model-building EA related to EDAs, requires multivariate linkage-learning to solve the problem in polynomial time. If only univariate learning is allowed (i.e. no linkage) P3 is shown to require exponential time to solve Concatenated Traps. Therefore in this example, multivariate modelling is essential.

Some more general insight into when linkages are essential or inessential can potentially be gained through Walsh function decomposition of binary functions. More research is needed into how robust Walsh coefficients are in terms of detecting essential linkage, particularly under function scaling.

The original question on precision remains open.

5 Panel discussions

5.1 Discussion Forum: Early Insights on Design and Analysis of EDAs

Duc-Cuong Dang (Universität Passau, DE), Timo Kötzing (Hasso-Plattner-Institut, Universität Potsdam, DE), Marcus Schmidbauer (Universität Passau, DE), and Andrew M. Sutton (University of Minnesota – Duluth, US)

The first part of this session was devoted to early theoretical approaches to the analysis of *Estimation-of-Distribution Algorithms* (EDAs). The second one gave a short introduction to the *Cross-Entropy Method* (CEM).

First Part: Early Theory on EDAs

Mühlenbein et. al. [1] consider additively decomposed functions, and point out that simple genetic algorithms fail when they are deceptive. The authors present conditions for such functions on which the Boltzmann distribution factors into a product over marginal distributions, making them tractable for univariate EDAs. They also demonstrate that when these conditions do not hold, the factorization is approximate. Further, the authors consider a purely selection-based EDA using the Boltzmann distribution, called *Boltzmann EDA* (BEDA). The authors prove that, for a given ϵ, BEDA on OneMax finds the optimum in N_ϵ (specified in the paper) generations with probability at least $1 - \epsilon$.

Mühlenbein et. al. [2] consider EDAs with infinite populations to solve continuous maximisation problems. They define that an EDA converges *globally* if the average fitness of the population converges to the globally optimal fitness value as time goes to infinity. The authors prove that certain purely selection-based EDAs converge globally. Further, they prove that this is also the case for the *Factorized Distribution Algorithm* (FDA) with proportional selection.

Second Part: Introduction to CEM

The idea of CEM [3] when applied to an optimisation problem is to view it as a problem of estimating the probability of sampling solutions above some quality threshold; typically this is a rare event and the technique of *importance sampling* provides an efficient solution. This is equivalent to minimising the *Kullback-Leibler metrics* between a family of distributions defined by some model and the ideal sampling distribution (hence the name cross-entropy) and analytically solving this minimisation problem produces a specific algorithm for the model, like in the case of the univariate model we obtain the *Univariate Marginal Distribution Algorithm* (UMDA) and then *Population-based Incremental Learning* (PBIL) via some smoothing process.

References

1 Heinz Mühlenbein, Thilo Mahnig, and Alberto O. Rodriguez. 1999. Schemata, Distributions and Graphical Models in Evolutionary Optimization. *Journal of Heuristics* (1999), 215–247. https://doi.org/10.1023/A:1009689913453
2 Qingfu Zhang, and Heinz Mühlenbein. 2004. On the convergence of a class of estimation of distribution algorithms. *IEEE Transactions on Evolutionary Computation* (2004), 127-136. https://doi.org/10.1109/TEVC.2003.820663

3 Pieter-Tjerk de Boer, Dirk P. Kroese, Shie Mannor, and Reuven Y. Rubinstein. 2005. A Tutorial on the Cross-Entropy Method. *Annals of Operations Research* (2005), 19-67. https://doi.org/10.1007/s10479-005-5724-z

6 Open problems

6.1 Sampling of Rare Fitness Values

Duc-Cuong Dang (Universität Passau, DE)

We discuss the algorithmic design problem of efficiently discovering all fitness values, thus introducing the notion of (expected) cover time instead of optimisation time. This design perspective is useful in some situations where the number of possible fitness values is relatively small and they are well connected.

A prominent algorithm that follows this idea is the so-called Fitness Frequency Assignment [2] which completely ignores the actual fitness values but focuses on the frequencies at which these values are encountered by the parent and offspring populations. We formulate this algorithm in its general form as the $(\mu+\lambda)$-FFA algorithm using unary unbiased variation operators and show some of its invariants and a conditional upper bound on the cover time. It remains an open question how to prove unconditional bounds for some functions such as OneMax or Jump, since experiments with fixed $\lambda = 1$ suggest that they lie between $\Omega(n^2)$ and $O(n^2 \log n)$ depending on μ. Finally, we introduce the notion of the fitness accessibility graph, and show that learning this graph on the fly can improve the cover time in some settings.

As a general question to the research community on EDAs, we want to know what information and statistics are useful to learn or model, then to exploit in order to minimise the cover time. Benjamin and Timo suggest a multi-objectivization approach: run GSEMO on $(f(x), -f(x))$ and this gives a cover time of $O(n^2 \log n)$ for OneMax and Jump. In a related work, Aisha and Timo have a paper [1] where they analyse Novelty Search on OneMinMax and show a non-trivial upper bound $O(n^2)$ for the expected optimisation time.

References

1 D. Antipov, T. Kötzing, A. Radhakrishnan. *Greedy Versus Curious Parent Selection for Multi-objective Evolutionary Algorithms.* PPSN (3) 2024: 86-101.
2 T. Weise, Z. Wu, X. Li, Y. Chen, and J. Lässig. *Frequency Fitness Assignment: Optimization Without Bias for Good Solutions Can Be Efficient.* IEEE Trans. Evol. Comput. 27(4): 980-992 (2023).

Participants

- Alistair Benford
University of Birmingham, GB
- Jasmin Brandt
Universität Paderborn, DE
- Josu Ceberio Uribe
University of the Basque Country – Donostia, ES
- Duc-Cuong Dang
Universität Passau, DE
- Benjamin Doerr
Ecole Polytechnique – Palaiseau, FR
- Carola Doerr
Sorbonne University – Paris, FR
- Martin Fyvie
The Robert Gordon University – Aberdeen, GB
- Nikolaus Hansen
INRIA Saclay – Palaiseau, FR
- Jose Ignacio Hidalgo
Complutense University of Madrid, ES
- Joshua D. Knowles
SLB Cambridge Research, GB
- Timo Kötzing
Hasso-Plattner-Institut, Universität Potsdam, DE
- Martin S. Krejca
Ecole Polytechnique – Palaiseau, FR
- Johannes Lengler
ETH Zürich, CH
- John McCall
The Robert Gordon University – Aberdeen, GB
- Aishwarya Radhakrishnan
Hasso-Plattner-Institut, Universität Potsdam, DE
- Franz Rothlauf
Universität Mainz, DE
- Roberto Santana
University of the Basque Country – Donostia, ES
- Valentino Santucci
University for Foreigners of Perugia, IT
- Marcus Schmidbauer
Universität Passau, DE
- Jonathan L. Shapiro
University of Manchester, GB
- Marta Rosa Soto Ortiz
CaixaBank Tech – Madrid, ES
- Andrew M. Sutton
University of Minnesota – Duluth, US
- Dirk Thierens
Utrecht University, NL
- Vanessa Volz
CWI – Amsterdam, NL
- Carsten Witt
Technical University of Denmark – Lyngby, DK