



DAGSTUHL REPORTS

Volume 14, Issue 1, January 2024

From Proofs to Computation in Geometric Logic and Generalizations (Dagstuhl Seminar 24021) <i>Ingo Blechschmidt, Hajime Ishihara, Peter M. Schuster, and Gabriele Buriola</i>	1
Fusing Causality, Reasoning, and Learning for Fault Management and Diagnosis (Dagstuhl Seminar 24031) <i>Alessandro Cimatti, Ingo Pill, and Alexander Diedrich</i>	25
Representation, Provenance, and Explanations in Database Theory and Logic (Dagstuhl Seminar 24032) <i>Pablo Barcelo, Pierre Bourhis, Stefan Mengel, and Sudeepa Roy</i>	49
Symmetric Cryptography (Dagstuhl Seminar 24041) <i>Christof Beierle, Bart Mennink, María Naya-Plasencia, Yu Sasaki, and Rachele Heim Boissier</i>	72
The Emerging Issues in Bioimaging AI Publications and Research (Dagstuhl Seminar 24042) <i>Jianxu Chen, Florian Jug, Susanne Rafelski, and Shanghang Zhang</i>	90
Next Generation Protocols for Heterogeneous Systems (Dagstuhl Seminar 24051) <i>Stephanie Balzer, Marco Carbone, Roland Kuhn and Peter Thiemann</i>	108
Reviewer No. 2: Old and New Problems in Peer Review (Dagstuhl Seminar 24052) <i>Iryna Gurevych, Anna Rogers, Nihar Shah, and Jingyan Wang</i>	130

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

July, 2024

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>.

License

This work is licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0).



In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:

- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

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.

Editorial Board

- Elisabeth André
- Franz Baader
- Goetz Graefe
- Reiner Hähnle
- Barbara Hammer
- Lynda Hardman
- Steve Kremer
- Rupak Majumdar
- Heiko Mantel
- Lennart Martens
- Albrecht Schmidt
- Wolfgang Schröder-Preikschat
- Raimund Seidel (*Editor-in-Chief*)
- Heike Wehrheim
- Verena Wolf
- Martina Zitterbart

Editorial Office

Michael Wagner (*Managing Editor*)
Michael Didas (*Managing Editor*)
Jutka Gasiorowski (*Editorial Assistance*)
Dagmar Glaser (*Editorial Assistance*)
Thomas Schillo (*Technical Assistance*)

Contact

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

Digital Object Identifier: 10.4230/DagRep.14.1.i

From Proofs to Computation in Geometric Logic and Generalizations

Ingo Blechschmidt^{*1}, Hajime Ishihara^{*2}, Peter M. Schuster^{*3}, and Gabriele Buriola^{†4}

1 Universität Augsburg, DE. iblech-dagstuhl@speicherleck.de

2 Toho University – Chiba, JP. hajime.ishihara@ext.toho-u.ac.jp

3 University of Verona, IT. petermichael.schuster@univr.it

4 University of Verona, IT. gabriele.buriola@univr.it

Abstract

What is the computational content of proofs? This is one of the main topics in mathematical logic, especially proof theory, that is of relevance for computer science. The well-known foundational solutions aim at rebuilding mathematics constructively almost from scratch, and include Bishop-style constructive mathematics and Martin-Löf's intuitionistic type theory, the latter most recently in the form of the so-called homotopy or univalent type theory put forward by Voevodsky.

From a more practical angle, however, the question rather is to which extent *any given* proof is effective, which proofs of which theorems can be rendered effective, and whether and how numerical information such as bounds and algorithms can be extracted from proofs. Ideally, all this is done by manipulating proofs mechanically and/or by adequate metatheorems (proof translations, automated theorem proving, program extraction from proofs, proof mining, etc.).

A crucial role for answering these questions is played by *coherent and geometric theories and their generalizations*: not only that they are fairly widespread in modern mathematics and non-classical logics (e.g., in abstract algebra, and in temporal and modal logics); those theories are also a priori amenable for constructivisation (see Barr's Theorem, especially its proof-theoretic variants, and the numerous Glivenko-style theorems); last but not least, effective theorem-proving for coherent theories can be automated with relative ease and clarity in relation to resolution.

Specific topics that substantially involve computer science research include categorical semantics for geometric theories up to the proof-theoretic presentation of sheaf models and higher toposes; extracting the computational content of proofs and dynamical methods in quadratic form theory; the interpretation of transfinite proof methods as latent computations; complexity issues of and algorithms for geometrization of theories; the use of geometric theories in constructive mathematics including finding algorithms, ideally with integrated developments; and coherent logic for obtaining automatically readable proofs.

Seminar January 7–12, 2024 – <https://www.dagstuhl.de/24021>

2012 ACM Subject Classification Theory of computation → Constructive mathematics; Theory of computation → Proof theory; Theory of computation → Automated reasoning

Keywords and phrases automated theorem proving, categorical semantics, constructivisation, geometric logic, proof theory

Digital Object Identifier 10.4230/DagRep.14.1.1

* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

From Proofs to Computation in Geometric Logic and Generalizations, *Dagstuhl Reports*, Vol. 14, Issue 1, pp. 1–24

Editors: Ingo Blechschmidt, Hajime Ishihara, and Peter M. Schuster



Dagstuhl Reports

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

1 Executive Summary

Ingo Blechschmidt (Universität Augsburg, DE)

Hajime Ishihara (Toho University – Chiba, JP)

Peter M. Schuster (University of Verona, IT)

License  Creative Commons BY 4.0 International license
© Ingo Blechschmidt, Hajime Ishihara, and Peter M. Schuster

The Dagstuhl Seminar 24021 emerged as a response to the challenges faced by its predecessor, Dagstuhl Seminar 21472, which grappled with pandemic-induced travel restrictions that hindered in-person attendance. The tireless efforts of the Dagstuhl staff notwithstanding, the earlier seminar relied on remote participation, limiting the depth of engagement and interaction among attendees. Many participants advocated for a follow-up seminar on a related topic, which materialized in the form of the present gathering.

Freed from the constraints of travel restrictions, this seminar hosted a dynamic environment characterized by extensive interactions, both structured and informal. Evening sessions provided a platform for casual discussions, enabling participants to delve deeper into topics of interest and forge meaningful connections across different communities within the field. Most notably, the seminar structure allowed ample time for spontaneous working groups.

As compared to the Dagstuhl Seminars the organisers attended in the past, this seminar stands out for an intense interaction between the participants. In the few months after we have already observed push effects to current research in several directions, e.g. strong negation in constructive mathematics and proof systems, synthetic algebraic geometry especially in the context of homotopy type theory, and topos theory. According to our humble opinion these effects can also be traced back to our stressing of interactive communication formats during the week in Dagstuhl, as there are the lightning talks sessions (one-hour slots of short talks of 5 minutes each) and especially the working groups, unusual both in number and participation.

As a pity, one of the four organisers, Thierry Coquand, could not attend this Dagstuhl Seminar. He still played a decisive role in forming the programme, especially concerning the crucial topics of geometric logic, topos theory and synthetic algebraic geometry with homotopy type theory.

An experimental addition to this year's seminar was the integration of an informal Discord server, serving as a digital hub for participant engagement. This platform not only facilitated pre-seminar planning and coordination but also granted participants a stronger sense of ownership over the seminar proceedings. Through features such as photo-sharing of blackboards and solicitation of talk topics, many participants actively shaped the agenda and direction of discussions.

2 Table of Contents

Executive Summary

<i>Ingo Blechschmidt, Hajime Ishihara, and Peter M. Schuster</i>	2
--	---

Overview of Talks

Geometric classes for the lazy topos theorist <i>Peter Arndt</i>	5
The Countable Reals <i>Andrej Bauer</i>	6
Quantified Boolean Formulas: Proofs, Solving and Circuits <i>Olaf Beyersdorff</i>	6
Skolem's Theorem in Coherent Logic <i>Marc Bezem</i>	7
A primer to realizability theory <i>Ingo Blechschmidt</i>	7
Geometric Type Theory <i>Ulrik Buchholtz</i>	8
Making predicative sense of impredicative foundations <i>Ulrik Buchholtz</i>	9
Introduction to synthetic algebraic geometry <i>Felix Cherubini and Matthias Hutzler</i>	9
Defining subsets of rational numbers using first-order logic <i>Nicolas Daans</i>	10
Hierarchy of Σ_n -fragments of logical principles over HA and hierarchy of intermediate propositional logics <i>Makoto Fujiwara</i>	11
Proof Mining: Logical Foundations and Recent Applications <i>Ulrich Kohlenbach</i>	11
Geometric theories from the point of view of constructive mathematics <i>Henri Lombardi</i>	12
Joyal's Arithmetic universes via the pure existential completion <i>Maria Emilia Maietti</i>	14
Structural proof theory for logics of strong negation <i>Sara Negri</i>	14
König's lemma, Weak König's lemma and Σ_1^0 induction <i>Takako Nemoto</i>	15
From a Constructive Logic to a Contradictory Logic <i>Satoru Niki</i>	16
Topologies of open complemented subsets <i>Iosif Petrakis</i>	16
Looking for the ideal background theory <i>Michael Rathjen</i>	17

A topos for continuous logic <i>Benno van den Berg</i>	17
Is geometric logic constructive? <i>Steven J. Vickers</i>	18
The Fundamental Theorem of Calculus, point-free <i>Steven J. Vickers</i>	18
Working groups	
Finiteness in synthetic algebraic geometry <i>Ingo Blechschmidt, Andrej Bauer, Felix Cherubini, Martín H. Escardó, and Matthias Hutzler</i>	19
Formalizing infinite time Turing machines <i>Ingo Blechschmidt, Andrej Bauer, Karim Johannes Becher, and Martín H. Escardó</i>	19
Tutorial on Agda, the dependently typed proof assistant <i>Ingo Blechschmidt</i>	20
Proof systems for geometric axioms and beyond <i>Sara Negri, Giulio Fellin, Eugenio Orlandelli, Edi Pavlovic, and Elaine Pimentel</i> .	20
Maps as bundles for point-free spaces <i>Steven J. Vickers</i>	21
Open problems	
On the status of Zorn’s lemma <i>Ingo Blechschmidt</i>	22
Remarks on predicativity <i>Stefan Neuwirth</i>	22
Participants	24

3 Overview of Talks

3.1 Geometric classes for the lazy topos theorist

Peter Arndt (Heinrich-Heine-Universität Düsseldorf, DE)

License © Creative Commons BY 4.0 International license
© Peter Arndt

Any geometric morphism between Grothendieck toposes factors as a surjection followed by a dense inclusion, followed by a closed inclusion. If the given geometric morphism f goes from the classifying topos $Set[T]$ of a geometric theory T to the classifying topos $Set[S]$ of a geometric theory S , then the two new toposes showing up in the factorization can be described as follows: The first one is the classifying topos of S' , the geometric theory consisting of all sequents that are valid in $f^*(M_S)$, the S -model in $Set[T]$ classified by f . And the second one is the classifying topos of S'' , the geometric theory consisting of S and all geometric sequents of the form $\varphi \vdash \perp$ that are valid in $f^*(M_S)$.

Computing the intermediate toposes can be very hard for general geometric theories. However, the following three observations point to a way of making the factorization usable for real mathematical situations:

1. The factorization still works for κ -geometric morphisms, and the intermediate toposes are the classifying toposes for the corresponding κ -geometric theories (i.e. geometric theories, but with infinitary conjunctions of size $< \kappa$)
2. Any accessible category is the category of Set-models of a κ -geometric *presheaf theory*.
3. Any accessible functor between accessible categories arises from an essential κ -geometric morphism between classifying presheaf toposes.

Thus, if we are willing to introduce the extra cardinal κ , we can place ourselves into the world of presheaf κ -toposes and essential κ -geometric morphisms. Here everything becomes rather easily computable:

Theorem: For an essential κ -geometric morphism between presheaf κ -toposes, induced by a functor $f: C \rightarrow D$ between the small index categories, the intermediate toposes of the factorization are the presheaf toposes on the full subcategory of D whose objects are in the essential image of f , and the full subcategory of D whose objects admit a morphism into the essential image of f , respectively.

Sample applications:



A. The theory of 3-colorable graphs can be axiomatized by adding to the theory of graphs all those geometric negations that are valid for the triangle graph. This can be seen by factorizing the geometric morphism from Set to the classifying topos of graphs, that chooses the triangle graph. The domain of the closed inclusion part is the classifying topos of 3-colorable graphs, because a graph is 3-colorable if and only if it admits a morphism to the triangle graph.

B. There are geometric sequents allowing to distinguish free finitely generated groups with different numbers of generators. This can be seen by considering the two geometric morphisms from Set to the classifying topos of groups that pick free groups on different numbers n, m (both ≥ 2) of generators. The surjection-inclusion factorizations yield the classifying toposes of the geometric theories of these free groups. By the theorem, they are the presheaf toposes over the full subcategories whose object are the free group with m , resp. n , generators. Explicit computation shows that their idempotent completions are non-equivalent, hence the presheaf toposes are non-equivalent, hence the classified geometric theories are different.

C. There are no negated κ -geometric formulas that are satisfied by all Special groups (a notion from quadratic form theory) coming from von Neumann-regular rings, but not by general special groups. This can be seen by considering the morphism from the classifying κ -topos of von Neumann-regular rings to the classifying κ -topos of special groups ($\kappa \geq \aleph_1$), and using the result that every special group admits a morphism the special group of a von Neumann-regular ring.

3.2 The Countable Reals

Andrej Bauer (University of Ljubljana, SI)

License  Creative Commons BY 4.0 International license
 Andrej Bauer

Joint work of Andrej Bauer, James E. Hanson

Main reference Andrej Bauer, James E. Hanson: “The Countable Reals”, CoRR, Vol. abs/2404.01256, 2024.

URL <https://arxiv.org/abs/2404.01256>

In 1874 Georg Cantor published a theorem stating that every sequence of reals is avoided by some real, thereby showing that the reals are not countable. Cantor’s proof uses classical logic. There are constructive proofs, although they all rely on the axiom of countable choice. Can the real numbers be shown uncountable without excluded middle and without the axiom of choice?

In this evening session, we have a tour of the recent result that, perhaps astoundingly, higher-order intuitionistic logic cannot show the reals to be uncountable. The result rests on the construction of a certain sequence of reals, shown to exist by Joseph S. Miller from University of Wisconsin–Madison, with a strong counter-diagonalization property, and a novel variant of realizability making use of oracles with a certain uniformity restriction.

3.3 Quantified Boolean Formulas: Proofs, Solving and Circuits

Olaf Beyersdorff (Friedrich-Schiller-Universität Jena, DE)

License  Creative Commons BY 4.0 International license
 Olaf Beyersdorff

Main reference Olaf Beyersdorff, Benjamin Böhm: “Understanding the Relative Strength of QBF CDCL Solvers and QBF Resolution”, Log. Methods Comput. Sci., Vol. 19(2), 2023.

URL [http://dx.doi.org/10.46298/LMCS-19\(2:2\)2023](http://dx.doi.org/10.46298/LMCS-19(2:2)2023)

This talk gives an overview on connections between proof complexity and algorithms for propositional satisfiability (SAT) and quantified Boolean formulas (QBF). We particularly cover the correspondence between conflict-driven clause learning algorithms (CDCL) for SAT and QBF to resolution proof systems. These connections enable the analysis of SAT/QBF algorithms through proof complexity. While in SAT there is a tight relation between propositional resolution and (non-deterministic) CDCL, the picture is more complex as we can show the incomparability of Q-Resolution and (non-deterministic) QCDCL. Nevertheless, lower bounds on QBF resolution systems yield run-time lower bounds for different QCDCL variants. Such lower bounds can be obtained via strong relations between QBF proof systems and circuit classes.

References

- 1 Olaf Beyersdorff. Proof complexity of quantified Boolean logic – a survey. In Marco Benini, Olaf Beyersdorff, Michael Rathjen, and Peter Schuster, editors, *Mathematics for Computation (M4C)*, pages 353–391. World Scientific, 2022.
- 2 Olaf Beyersdorff, Joshua Blinkhorn, Meena Mahajan, and Tomás Peitl. Hardness characterisations and size-width lower bounds for QBF resolution. *ACM Transactions on Computational Logic*, 2022.
- 3 Knot Pipatsrisawat and Adnan Darwiche. On the power of clause-learning SAT solvers as resolution engines. *Artif. Intell.*, 175(2):512–525, 2011.

3.4 Skolem’s Theorem in Coherent Logic

Marc Bezem (*University of Bergen, NO*)

License © Creative Commons BY 4.0 International license
© Marc Bezem

Joint work of Marc Bezem, Thierry Coquand

Main reference Marc Bezem, Thierry Coquand: “Skolem’s Theorem in Coherent Logic”, *Fundam. Informaticae*, Vol. 170(1-3), pp. 1–14, 2019.

URL <http://dx.doi.org/10.3233/FI-2019-1853>

After a brief introduction to skolemization and to coherent logic, we discuss three possible research directions, relevant for automated reasoning.

- Generalization of skolemization from (properly quantified) atoms to coherent sentences. Due to the implication in the latter, one could call this conditional (or hypothetical) skolemization.
- More efficient proof systems for coherent logic. Ground forward reasoning is conceptually very simple and therefore attractive, but it leads to exponentially longer proofs as compared to non-ground rules.
- Comparison of the length of skolemized and non-skolemized proofs. The latter are known to be (much) longer in the case of first-order logic, but this is unknown for the coherent fragment.

3.5 A primer to realizability theory

Ingo Blechschmidt (*Universität Augsburg, DE*)

License © Creative Commons BY 4.0 International license
© Ingo Blechschmidt

Main reference Andrej Bauer: “Realizability as the connection between computable and constructive mathematics”, 2005.

URL <https://math.andrej.com/asset/data/c2c.pdf>

Realizability theory provides a bridge between constructive and computable mathematics. In its basic form, the fundamental result is that there is a mechanical way to extract from any given constructive number-theoretic proof a computable witness, a so-called “realizer” for instance, a realizer for the infinitude of primes is a Turing machine computing larger and larger prime numbers.

From this basic picture, originally due to Kleene, gradually several enhancements emerged: to higher-order statements and proofs; to proofs set in flavors of classical set theory; and to different machine models.

This evening talk featured a leisurely introduction to realizability theory. For building intuition, we focused on the following examples, exploring for each its status in classical mathematics, constructive mathematics, ordinary realizability and realizability based on infinite time Turing machines:

1. Every number is prime or not prime.
2. Every map $\mathbb{N} \rightarrow \mathbb{N}$ has a zero or not.
3. Every map $\mathbb{N} \rightarrow \mathbb{N}$ is (Turing-)computable.
4. Every map $\mathbb{R} \rightarrow \mathbb{R}$ is continuous.
5. Markov's principle holds.
6. Countable choice holds.
7. Heyting arithmetic is categorical.
8. A statement holds iff it is realized.
9. There is an injection $\mathbb{R} \hookrightarrow \mathbb{N}$.

Our decision to approach our exploration in a constructive metatheory allowed us to discuss inheritance of classical principles. For instance, Markov's principle is realized if and only if it holds in the metatheory, while countable choice is realized even if it fails in the metatheory.

3.6 Geometric Type Theory

Ulrik Buchholtz (University of Nottingham, GB)

License  Creative Commons BY 4.0 International license
© Ulrik Buchholtz

This was an exploratory talk, outlining the structure of an envisioned geometric type theory.

The idea is to have a type theory capable of formalizing geometric reasoning, proving results such as $((\phi \rightarrow \perp) \rightarrow \perp) = 1$ for the generic proposition ϕ , and similarly other non-geometric properties of generic objects; and also capable of constructing common geometric maps and bundles.

The intended model is based on (small) toposes (or $(\infty, 1)$ -toposes for a homotopical version), where

- contexts are such toposes \mathcal{X}

and with two different type judgments:

- $\mathcal{X} \vdash A \text{ étale}_i$ corresponding to sheaves over \mathcal{X} in \mathcal{U}_i (classified by the usual univalent universes \mathcal{U}_i)
- $\mathcal{X} \vdash Y \text{ space}_i$ corresponding to geometric morphisms $p : \mathcal{Y} \rightarrow \mathcal{X}$ in \mathcal{U}_i .

We mentioned that maybe we should allow Y itself to be in \mathcal{U}_j , necessitating two indices, $\text{space}_{i,j}$.

Correspondingly, there should be two different element judgments:

- $\mathcal{X} \vdash a : A$ corresponding to sections of sheaves,
- $\mathcal{X} \vdash y : Pt_i(Y)$ corresponding to geometric sections $y : \mathcal{X} \rightarrow \mathcal{Y}$ in \mathcal{U}_j .

We also need to consider two fragments:

- A geometric fragment ($\Sigma, \text{Id}, \text{finitary HITs}$) stable under all geometric morphisms, and
- A full fragment ($\Pi, \text{W}, \text{universes}$) stable only under étale morphisms.

An inspiration is Mike Shulman's observation that in the $(\infty, 1)$ -world atomic=logical=étale.

A question arose as to whether we prove that spaces are exponentiable, or whether we need a separate judgment for that. If we write $[Y, Z]$ for the exponential, then if A, B are étale, the Π -type $A \rightarrow B$ should be the discrete coreflection $(A \rightarrow B) \rightarrow [A, B]$.

For the generic proposition (and hence all propositions) ϕ , we should get $[[\phi, 0], 0] = \phi$.

3.7 Making predicative sense of impredicative foundations

Ulrik Buchholtz (*University of Nottingham, GB*)

License © Creative Commons BY 4.0 International license
© Ulrik Buchholtz

Main reference Solomon Feferman: “Does reductive proof theory have a viable rationale?” *Erkenntnis* 53: 63–96, 2000.

URL <https://doi.org/10.1023/A:1005622403850>

Although classical principles such as the law of excluded middle or the axiom of choice are not available in constructive mathematics, constructive mathematics can make sense of them: The double-negation translation interprets classical logic in intuitionistic logic, and Gödel’s L translation embeds ZFC in ZF. In this way, proofs in classical mathematics can be understood as blueprints for constructive proofs; classical proofs can be decrypted to unearth obscured constructive content. Several other techniques for making constructive sense of classical proofs exist as well. Adding the law of excluded middle or the axiom of choice typically does not increase the consistency strength at all.

In contrast, there is no similar translation from impredicative mathematics to predicative mathematics; indeed, impredicative foundations are typically much stronger than their predicative counterparts. Hence it seems that predicative foundations cannot make sense of impredicative systems.

The evening talk explored that, fantastically, in some cases it is still possible to extract predicative content from impredicative proofs. As direct translations are no longer possible, the talk offered a whirlwind tour of the required tool, ordinal analysis, including a discussion of the true limits of predicative reasoning.

3.8 Introduction to synthetic algebraic geometry

Felix Cherubini (*Chalmers University of Technology – Göteborg, SE*) and Matthias Hutzler (*University of Gothenburg, SE*)

License © Creative Commons BY 4.0 International license
© Felix Cherubini and Matthias Hutzler

Joint work of Felix Cherubini, Thierry Coquand, Matthias Hutzler

Algebraic Geometry is about studying shapes given by polynomial equations. Systems of polynomial equations are mapped to some “space” of solutions (the common zeros). It turns out that we can just use sets as “spaces” if we do not insist on the axiom of choice or the law of excluded middle. In their absence we can ask that the SQC axiom (of Blechschmidt) holds: The map $A \rightarrow R^{\text{Spec}(A)}$ is an isomorphism for all finitely presented R -algebras $A = R[X_1, \dots, X_n]/(P_1, \dots, P_\ell)$ and $\text{Spec}(A) := \{x \in R^n \mid \forall i. P_i(x) = 0\}$. Together with two other axioms, this is the start of a well-working synthetic version of algebraic geometry which is modeled by a sheaf topos, the Zariski topos.

In the talk we also mentioned the topic of finite schemes in the synthetic setting, hoping to spark new ideas in connecting the geometric-algebraic notion of a finite scheme with finiteness notions in constructive mathematics.

References

- 1 Felix Cherubini, Thierry Coquand, Matthias Hutzler. *A Foundation for Synthetic Algebraic Geometry*. arXiv:2307.00073, 2023

3.9 Defining subsets of rational numbers using first-order logic

Nicolas Daans (Charles University – Prague, CZ)

License  Creative Commons BY 4.0 International license
© Nicolas Daans

Joint work of Becher, Karim Johannes; Daans, Nicolas; Dittmann, Philip; Fehm, Arno
Main reference Nicolas Daans: “Existential first-order definitions and quadratic forms”. Phd thesis. Universiteit Antwerpen, 2022. Under the supervision of Karim Johannes Becher and Philip Dittmann.
URL <https://hdl.handle.net/10067/1903760151162165141>

In 1970, building on work of Davis, Putnam, and Robinson, Matiyasevich established a complete classification of subsets of the integers which can be described by an existential first-order formula in the language of rings, in terms of their computability by a Turing machine. After Gödel’s (in)famous Incompleteness Theorems, this marks one of the major milestones in the study of computability in number theory, and it has inspired continued research into (existential) first-order definability of subsets of rings.


This evening talk offers a lighthearted initiation into the realm of first-order and existential definability within fields, with a focus on the field of rational numbers. Through various examples, I highlight the intricacies surrounding the classification of definable subsets of the rationals and associated decidability problems, while showcasing the diverse array of methodologies drawn from algebra, number theory, quadratic form theory, and model theory that have been employed to tackle them. To align with the theme of this week’s seminar, particular attention is devoted to contrasting constructive and non-constructive approaches to these problems. Nonetheless, the talk remains accessible to audiences with classical training and assumes no prior background knowledge.

References

- 1 Nicolas Daans, Philip Dittmann, and Arno Fehm. “*Existential rank and essential dimension of diophantine sets*”. Available as arXiv:2102.06941. Oct. 2021. <https://doi.org/10.48550/arXiv.2102.06941>
- 2 Nicolas Daans. “*Existential first-order definitions and quadratic forms*”. Phd thesis. Universiteit Antwerpen, 2022. <https://hdl.handle.net/10067/1903760151162165141>
- 3 Nicolas Daans. “*Universally defining Z in Q with 10 quantifiers*”. In: Journal of the London Mathematical Society 109.2 (2024), e12864. <https://doi.org/10.1112/jlms.12864>
- 4 Yuri R. Matiyasevich. “*Diofantovost’ perechislimykh mnozhestv [Enumerable sets are Diophantine]*”. In: Doklady Akademii Nauk SSSR 191 (1970), pp. 279-282. <http://mi.mathnet.ru/dan35274>
- 5 Julia Robinson. “*Definability and decision problems in arithmetic*”. In: Journal of Symbolic Logic 14 (1949), pp. 98-114.

3.10 Hierarchy of Σ_n -fragments of logical principles over HA and hierarchy of intermediate propositional logics

Makoto Fujiwara (Tokyo University of Science, JP)

License  Creative Commons BY 4.0 International license
© Makoto Fujiwara

Joint work of Makoto Fujiwara, Hajime Ishihara, Takako Nemoto, Nobu-Yuki Suzuki, Keita Yokoyama
Main reference Makoto Fujiwara, Hajime Ishihara, Takako Nemoto, Nobu-Yuki Suzuki, Keita Yokoyama: “Extended Frames and Separations of Logical Principles”, Bull. Symb. Log., Vol. 29(3), pp. 311–353, 2023.
URL <http://dx.doi.org/10.1017/BSL.2023.29>

We investigate seven natural variations of the linearity axiom

$$\text{LIN} : (p \rightarrow q) \vee (q \rightarrow p)$$

and the following principles in the context of intermediate propositional logic and arithmetic:

- PEM (the principle of excluded middle) : $p \vee \neg p$;
- WPEM (the weak principle of excluded middle) : $\neg p \vee \neg\neg p$;
- DML (the De Morgan law) : $\neg(p \wedge q) \rightarrow \neg p \vee \neg q$;
- WDML (the weak De Morgan law) : $\neg(\neg p \wedge \neg q) \rightarrow \neg\neg p \vee \neg\neg q$;
- DNE (the double negation elimination) : $\neg\neg p \rightarrow p$.


In fact, the hierarchy of Σ_n -fragments of the logical principles over Heyting arithmetic HA is quite different from that of the intermediate propositional logics induced by the logical principles. For a separation result on the Σ_n -fragment of

$$\text{LIN}_7 : (p \rightarrow \neg\neg q) \vee (\neg\neg q \rightarrow p)$$

in arithmetic, we employ a meta-theorem with respect to extended frame to an appropriate propositional Kripke model which refutes LIN_7 .

3.11 Proof Mining: Logical Foundations and Recent Applications

Ulrich Kohlenbach (TU Darmstadt, DE)

License  Creative Commons BY 4.0 International license
© Ulrich Kohlenbach

In the Proof Mining paradigm one applies proof-theoretic transformations to extract new computational information (e.g. programs or effective bounds) as well as qualitative improvements (e.g. by weakening of assumptions) from given *prima facie* noneffective proofs in different areas of core mathematics. In this talk, we will

- discuss recent extensions (mainly due to [8] but see also [5]) of the existing framework for logical bound extraction metatheorems to cover set-valued monotone and accretive operators in Hilbert and Banach spaces;
- indicate the potential of generalizing mined proofs such as the quantitative analysis of a Halpern-type proximal point algorithm in suitable Banach spaces given in [3] to new geodesic settings ([7, 10]) as well as to the context of so-called Bregman strongly nonexpansive mappings ([8], [9]);
- show how a rate of convergence for the Dykstra algorithm in convex optimization which recently has been obtained in [6] under a metric regularity assumption can be explained in terms of a novel extension of the concept of Fejér monotonicity ([4]);

- prove the first linear rates of asymptotic regularity for the Tikhonov-Mann algorithm in geodesic spaces as an application of proof mining ([1]);
- present the first effective rate of convergence for the asymptotic regularity of ergodic averages of iterations of nonexpansive mappings in uniformly convex Banach spaces ([2]).

References

- 1 H. Cheval, U. Kohlenbach, L. Leuştean: On modified Halpern and Tikhonov-Mann iterations. *Journal of Optimization Theory and Applications* vol.197, pp. 233-251 (2023).
- 2 A. Freund, U. Kohlenbach: R.E. Bruck, proof mining and a rate of asymptotic regularity for ergodic averages in Banach spaces. *Applied Set-Valued Analysis and Optimization* vol.4, pp. 323-336 (2022).
- 3 U. Kohlenbach: Quantitative analysis of a Halpern-type proximal point algorithm for accretive operators in Banach spaces. *Journal of Nonlinear and Convex Analysis* vol. 21, no.9, pp. 2125-2138 (2020).
- 4 U. Kohlenbach, P. Pinto: Fejér monotone sequences revisited. Submitted
- 5 U. Kohlenbach, N. Pischke: Proof theory and nonsmooth analysis. *Philosophical Transactions of the Royal Society A* vol.381, Issue 2248, DOI 10.1098/rsta.2022.0015, 21 pages (2023).
- 6 P. Pinto: On the finitary content of Dykstra's cyclic projections algorithm. arXiv:2306.09791, 2023, submitted.
- 7 P. Pinto: Nonexpansive maps in nonlinear smooth spaces. Submitted.
- 8 N. Pischke: Proof-Theoretical Aspects of Nonlinear and Set-Valued Analysis. PhD Thesis, Technische Universität Darmstadt, xvi+338pp., 2024.
- 9 N. Pischke, U. Kohlenbach: Effective rates for iterations involving Bregman strongly nonexpansive mappings. 2024 Submitted.
- 10 A. Sipoş: Abstract strongly convergent variants of the proximal point algorithm. *Computational Optimization and Applications* vol.83, pp. 349-380 (2022).

3.12 Geometric theories from the point of view of constructive mathematics

Henri Lombardi (University of Franche-Comté – Besançon, FR)

License  Creative Commons BY 4.0 International license
© Henri Lombardi

1) Constructive mathematics.

Gauss, Kronecker, Poincaré, Bishop.

Intuitive mathematics, (no need of formalisation) using basic, undefined tools :

- natural numbers, constructions,
- proofs : a proof is a convincing proof,
- predicativity,
- Poincaré: Never lose sight of the fact that every proposition concerning infinity must be the translation, the precise statement of propositions concerning the finite.

2) Why using geometric theories ?

Formalisation is about a piece of intuitive informal mathematics. It can help to understand and analyse this piece of mathematics.

Geometric theories are interesting because they use only very simple kinds of assertions, geometric assertions, managed without using logic: they are purely computational machineries.

Precursor: recursive arithmetic by Goodstein.

My understanding :

- sorts in a geometric theories correspond to sets in Bishop constructive mathematics (BCM)
- geometric theories are to be used and analysed in the external world given by BCM

This is very different of the usual understanding of geometric theories where the external world is a world of classical mathematics, using categorical logic, LEM and Choice, perhaps also power sets and other ugly things à la ZFC.

The underlying logic of geometric theories is a small part of minimal logic and natural deduction: no negation, no connector \Rightarrow , no complicated sentences using forall exists forall.

No conflict between classical and constructive understanding of mathematics used inside geometric theories. Adding formally negation and \Rightarrow does not add new valid rules in the former language.

3) Dynamic versus geometric theories.

The dynamic view of geometric theories is purely computational, logic free

- no new formulas: only atomic formulas given in the language
- no logical implication: only deduction rules
- connector “or” is replaced by “open branches of computations”
- quantification “exists y such that $P(y)$ ” is replaced by “introduce a new fresh variable y satisfying $P(y)$ ”
- a deduction rule is valid when you have constructed a deduction tree:
 - at the root you have your context and your hypothesis,
 - at each node you use axioms as previously explained,
 - and the conclusion is to be valid at each leaf.

4) Strength of geometric theories.

A large part of usual abstract algebra can be formalised in **finitary geometric theories**, which corresponds to usual coherent first order theories in logic.

E.g., groups, lattice groups, local rings, residually discrete local rings, Prüfer domains, finitely presented modules, gcd rings, f-rings, discrete fields, discrete ordered fields, real closed discrete fields, discrete valued fields, henselian fields, ...

Some concepts need **infinitary geometric theories**. E.g. flatness, Krull dimension, coherence, Dedekind domains, Krull rings, projectively resolvable modules, ...

Some concepts, as Noetherianity, are outside the scope of geometric theory.

Infinitary geometric theories allow infinite disjunctions in the conclusion of a dynamic rule. The external world becomes crucial when constructing deduction trees.

5) Grothendieck toposes.

In classical mathematics, Grothendieck toposes are more or less “the same objects” as geometric theories. Grothendieck coherent toposes correspond to finitary geometric theories.

This important fact has to be understood in the context of constructive mathematics with an external world à la BCM.

Particularly important is to interpret geometric functors between Grothendieck toposes as corresponding to some kinds of morphisms between geometric theories, with an external world à la BCM.

3.13 Joyal’s Arithmetic universes via the pure existential completion

Maria Emilia Maietti (University of Padova, IT)

License  Creative Commons BY 4.0 International license
© Maria Emilia Maietti

Joint work of Maria Emilia Maietti, Davide Trotta

Arithmetic universes were built by A. Joyal in the seventies in some lectures, all still unpublished, to provide a categorical proof of Gödel’s incompleteness theorems. They amount to be exact completions of the lex category of predicates of a given Skolem theory. In [1] we proposed the notion of list-arithmetic pretopos as the general abstract notion of arithmetic universe.


In our talk, we report recent results obtained in joint work with Davide Trotta in [3]. We show that each arithmetic universe is the exact completion of the pure existential completion of the doctrine of its Skolem predicates. The notion of existential completion was introduced by D. Trotta in [4] and further analyzed in [2] in terms of existential free objects relative to an existential doctrine. In particular, we show in [3] that the Skolem predicates of an arithmetic universe provide a projective cover of the exact completion of their free pure existential completion. As a byproduct, we conclude that the initial arithmetic universe is the exact completion of the doctrine of recursively enumerable predicates.

References

- 1 M.E. Maietti. Joyal’s arithmetic universe as list-arithmetic pretopos. *Theory Appl. Categ.*, 24(3):39–83, 2010
- 2 M.E. Maietti and D. Trotta. A characterization of generalized existential completions. *Ann. Pure Appl. Logic*, 174(4):103234, 2023.
- 3 M.E. Maietti and D. Trotta. A characterization of regular and exact completions of pure existential completions. *Arxiv*, <https://arxiv.org/abs/2306.13610>, 2023
- 4 D. Trotta. The existential completion. *Theory Appl. Categ.*, 35(43):1576–1607, 2020.

3.14 Structural proof theory for logics of strong negation

Sara Negri (University of Genova, IT)

License  Creative Commons BY 4.0 International license
© Sara Negri

Joint work of Norihiro Kamide, Sara Negri

Gurevich logic is an extended constructive three-valued logic obtained from intuitionistic logic by adding a connective \sim of *strong negation*, with the following axiom schemata, where \neg is intuitionistic negation:

1. $\sim\sim A \supset A$,
2. $\sim\neg A \supset A$,
3. $\sim A \supset \neg A$,
4. $\sim(A \wedge B) \supset \sim A \vee \sim B$,
5. $\sim(A \vee B) \supset \sim A \wedge \sim B$,
6. $\sim(A \supset B) \supset A \wedge \sim B$.

Nelson logic, also known as Nelson’s constructive three-valued logic N3, is the intuitionistic negation-less fragment of Gurevich logic.

The primary formal difficulty in developing a natural deduction system for logics of strong negation lies in the requirement of having rules for \neg and \sim without \perp . This is solved using the rules of *explosion*, of \neg -*introduction*, and of *excluded middle*:

$$\frac{\neg A \quad A}{C} \text{Exp} \qquad \frac{\begin{array}{c} [A] \\ \vdots \\ C \end{array} \quad \begin{array}{c} [A] \\ \vdots \\ \neg C \end{array}}{\neg A} \neg\text{I} \qquad \frac{\begin{array}{c} [\neg A] \\ \vdots \\ C \end{array} \quad \begin{array}{c} [A] \\ \vdots \\ C \end{array}}{C} \text{Em}$$

The following are presented:

- Natural deduction and G3-style sequent calculi systems for Gurevich and Nelson logic.
- A translation between the natural deduction and sequent calculi introduced, with an indirect proof of normalization as a consequence of cut elimination.
- A syntactic embedding of Gurevich logic into intuitionistic logic.
- A classical sequent calculus with strong negation, used as a platform for obtaining classical logics of strong negation, such as Avron and De-Omori logics.
- A Glivenko theorem for embedding classical logic with strong negation into Gurevich logic.

References

- 1 N. Kamide and S. Negri, Unified natural deduction for logics of strong negation, ms., 2023.
- 2 N. Kamide and S. Negri, G3-style sequent calculi for Gurevich logic and its neighbors, ms., 2023.

3.15 König's lemma, Weak König's lemma and Σ_1^0 induction

Takako Nemoto (Tohoku University – Sendai, JP)

License  Creative Commons BY 4.0 International license
© Takako Nemoto

Joint work of Takako Nemoto, Makoto Fujiwara

It is known that weak König's lemma (WKL) is strictly weaker than König's lemma (KL) in Friedman-Simpson's reverse mathematics ([2]). Then a natural question is how we can characterise the difference between WKL and KL. In Simpson's book, it is stated:

- WKL does not imply Σ_1^0 induction.
- WKL is equivalent to KL for trees with height-wise bounded functions over RCA_0 .

It is also known that Fan theorem, a classical contraposition of KL, yields Σ_1^0 induction over RCA_0^* ([1]). In this talk, we consider the difference between WKL and KL in the context of constructive reverse mathematics and the relationship between KL and Σ_1^0 induction.

References

- 1 T. Nemoto and K. Sato, *A marriage of Brouwer's Intuitionism and Hilbert's Finitism I: Arithmetic*, The Journal of Symbolic Logic, 87 (2), pp.437-497 (2022) doi:10.1017/jsl.2018.6
- 2 Subsystems of second order arithmetic, CUP (2009)

3.16 From a Constructive Logic to a Contradictory Logic

Satoru Niki (Ruhr-Universität Bochum, DE)

License  Creative Commons BY 4.0 International license
© Satoru Niki

The negation in intuitionistic logic has sometimes been criticised as problematic. One reason for this is the alleged lack of witness for negated conjunctions: $\text{not}-(A \text{ and } B)$ may hold without one of $\text{not}-A$ and $\text{not}-B$ being so. As a remedy, David Nelson introduced an alternative notion of “constructible falsity”, which promises such a witness. Nelson’s negation can even be made paraconsistent, which also encouraged the abandonment of intuitionistic negation altogether. However, such a move may obscure the notion of negation from the viewpoint of an intuitionistic logician. In this talk, I am going to give an overview of some different views for constructive negation. Then I will suggest another solution to the problem of negated conjunction. This solution, which leads to a variant of Heinrich Wansing’s logic C, hints that intuitionistic negation and paraconsistent constructible falsity may be seen as complementary, once we accept some contradictions to be provable.

3.17 Topologies of open complemented subsets

Iosif Petrakis (University of Verona, IT)

License  Creative Commons BY 4.0 International license
© Iosif Petrakis

We present cs-topologies, or topologies of complemented subsets, as a new approach to constructive topology that preserves the duality between open and closed subsets of classical topology. Complemented subsets were used successfully by Bishop in his constructive formulation of the Daniell approach to measure and integration. We use complemented subsets in topology, in order to describe simultaneously an open set, the first-component of an open complemented subset, and its complement as a closed set, the second component of an open complemented subset. We analyse the canonical cs-topology induced by a metric, and we introduce the notion of a modulus of openness for a cs-open subset of a metric space. Pointwise and uniform continuity of functions between metric spaces are formulated with respect to the way these functions inverse open complemented subsets together with their moduli of openness. The addition of moduli of openness in the concept of an open subset, given a base for the cs-topology, makes possible to define the notion of uniform continuity of functions between csb-spaces, that is cs-spaces with a given base. In this way, the notions of pointwise and uniform continuity of functions between metric spaces are directly generalised to the notions of pointwise and uniform continuity between csb-topological spaces.

References

- 1 I. Petrakis. *Topologies of open complemented subsets*. arXiv:2312.17095v1, 2023.

3.18 Looking for the ideal background theory

Michael Rathjen (University of Leeds, GB)

License © Creative Commons BY 4.0 International license
© Michael Rathjen

Proofs in mathematics often have a narrative quality to them, taking the reader on a long journey. Sometimes the reader has to wait for new mathematical characters (like imaginary numbers) to be created so the journey can be continued. Hilbert called these novel characters ideal elements. His conservation program was the idea that, while being important for the advancement of mathematics, ideal elements should be eliminable from proofs of concrete mathematical theorems.

Investigations by a long list of mathematicians/logicians (e.g. Weyl, Hilbert, Bernays, Lorenzen, Takeuti, Feferman, Friedman, Simpson to name a few) have shown that large swathes of ordinary mathematics can be undergirded by theories of fairly modest consistency strength. This confirms what Hilbert surmised in his program, namely that elementary results (e.g. those expressible in the language of number theory) proved in abstract, non-constructive mathematics can often be proved by elementary means.

The best known program for calibrating the strength of theorems from ordinary mathematics is reverse mathematics (RM). RM's scale for measuring strength is furnished by certain standard systems couched in the language of second order arithmetic. However, its language is not expressive enough to be able to talk about higher order objects, such as function spaces, directly.

Richer formal systems, in which higher order mathematical objects can be directly accounted for, have been suggested. The price for maintaining conservativity over elementary theories, however, is that one has to use different logics for different ontological realms, allowing classical logic to reign at the level of numbers whereas higher type mathematical objects are merely answerable to intuitionistic logic. In the talk, I'd like to present some of these semi-intuitionistic systems, give a feel for carrying out mathematics within them, and relate them to systems considered in RM.

3.19 A topos for continuous logic

Benno van den Berg (University of Amsterdam, NL)

License © Creative Commons BY 4.0 International license
© Benno van den Berg

Joint work of Daniel Figueroa

Main reference Daniel Figueroa, Benno van den Berg: "A topos for continuous logic". *Theory Appl. Categ.* 38, 1108-1135 (2022)

URL <http://www.tac.mta.ca/tac/volumes/38/28/38-28.pdf>

Continuous model theory is an area of model theory in which one studies structures where the truth values of sentences can be any element of the unit interval. This branch of model theory has applications in areas where metric structures are pervasive, such as in analysis and probability theory.

In this talk we propose an analysis of continuous model theory based on ideas from categorical logic, in particular Lawvere's notion of a hyperdoctrine. We propose a hyperdoctrine for continuous model theory and show that this hyperdoctrine can be embedded in the subobject hyperdoctrine of a topos. This means that continuous logic can be understood as the internal logic of a suitable topos.

3.20 Is geometric logic constructive?

Steven J. Vickers (*University of Birmingham, GB*)

License  Creative Commons BY 4.0 International license
© Steven J. Vickers

Joint work of Ming NG

For point-free topology (locales, etc.) we can reason validly in a natural manner using points, provided we restrict ourselves to constructions that are geometric – preserved by pullback along geometric morphisms. This style has now been applied in some results of real analysis, including exponentiation and logarithms (joint with Ming Ng) and the Fundamental Theorem of Calculus.

Are such constructions constructive in the strong sense of yielding algorithms? This is very unclear. My talk discusses some aspects of the question.

1. Some constructive taboos are true geometrically! But that is because they have to be interpreted topologically.
2. Geometric maths can be understood via an ontology of observations rather than constructions.
3. Point-free surjections, which embody conservativity principles, allow us to reason as if certain constructions can be done, even when they can't.

3.21 The Fundamental Theorem of Calculus, point-free

Steven J. Vickers (*University of Birmingham, GB*)

License  Creative Commons BY 4.0 International license
© Steven J. Vickers

Main reference Steven Vickers. *The Fundamental Theorem of Calculus point-free, with applications to exponentials and logarithms*. arXiv:2312.05228, 2023

URL <https://arxiv.org/abs/2312.05228>

The Fundamental Theorem of Calculus is so fundamental that, once it has been proved point-free, much subsequent development can be more or less standard. In a 5 minute minitalk I summarized the main features in [1] by which this was achieved.

1. Lower and upper integrals, valued as lower and upper reals, were described in [2]. After this, it is necessary only to show how they can be combined to give Dedekind reals.
2. Differentiation is treated in a Carathéodory style, with limits replaced by the existence of continuous slope maps. This was used in [3], where also Rolle's Theorem was proved.
3. For the derivative of an integral, the slope map can be defined as an integral with respect to the *uniform* valuation on $[x_0, x]$. Geometricity guarantees that this is continuous even at the limiting case $x = x_0$.

References

- 1 Steven Vickers. *The Fundamental Theorem of Calculus point-free, with applications to exponentials and logarithms*. arXiv:2312.05228, 2023
- 2 Steven Vickers. *A localic theory of lower and upper integrals*. *Mathematical Logic Quarterly* **54**, 2008
- 3 Steven Vickers. *The connected Vietoris powerlocale*. *Topology and its Applications* **156**, 2009

4 Working groups

4.1 Finiteness in synthetic algebraic geometry

Ingo Blechschmidt (Universität Augsburg, DE), Andrej Bauer (University of Ljubljana, SI), Felix Cherubini (Chalmers University of Technology – Göteborg, SE), Martín H. Escardó (University of Birmingham, GB), and Matthias Hutzler (University of Gothenburg, SE)

License © Creative Commons BY 4.0 International license
 © Ingo Blechschmidt, Andrej Bauer, Felix Cherubini, Martín H. Escardó, and Matthias Hutzler
Joint work of Ingo Blechschmidt, Andrej, Felix Cherubini, Martín H. Escardó, Hugo Moeneclaey, Matthias Hutzler, David Wärn
Main reference Ingo Blechschmidt, Felix Cherubini, Hugo Moeneclaey, Matthias Hutzler, David Wärn: “Finite Schemes in Synthetic Algebraic Geometry”, 2024
URL <https://felix-cherubini.de/finite.pdf>

In the related talk of Cherubini and Hutzler, the recent foundations for synthetic algebraic geometry were discussed [1]. One important step in improving synthetic calculations is understanding what notions of finiteness can help to move from geometric problems to numbers of interest. Ingo Blechschmidt presented a proof that finite schemes satisfy the finiteness notion of “Noetharian” from constructive mathematics [1]. This notion captures the intuition that there is no infinite sequence in a set. However, we found out in the group, that there are Noetherian sets in our language, which are not finite schemes. We discussed results on compactness of finite schemes which are currently expanded in [2] and Ingo Blechschmidt found a candidate for the synthetic version of the notion of “quasi-finite” from classical algebraic geometry. In a discussion with Martín Escardó and Andrej Bauer we checked if it is possible that the base ring of our theory is overt (a notion from synthetic topology). This turned out to be true, for example in the case where the external base ring is an algebra over an infinite field. To our surprise, overtness has a strong implications: non-empty open subsets of the base ring are inhabited.

References

- 1 T. Coquand, A. Spiwack. Constructively Finite?
- 2 I. Blechschmidt, F. Cherubini, H. Moeneclaey, M. Hutzler, D. Wärn. *Finite Schemes in Synthetic Algebraic Geometry*.
- 3 Felix Cherubini, Thierry Coquand, Matthias Hutzler. *A Foundation for Synthetic Algebraic Geometry*. arXiv:2307.00073, 2023

4.2 Formalizing infinite time Turing machines

Ingo Blechschmidt (Universität Augsburg, DE), Andrej Bauer (University of Ljubljana, SI), Karim Johannes Becher (University of Antwerp, BE), and Martín H. Escardó (University of Birmingham, GB)

License © Creative Commons BY 4.0 International license
 © Ingo Blechschmidt, Andrej Bauer, Karim Johannes Becher, and Martín H. Escardó
URL <https://agdapad.quasicoherent.io/Dagstuhl2024/html/ittm.html>

Infinite time Turing machines provide an exotic model of computation where the operation of Turing machines is extended to infinite ordinal time. They have been devised by Joel David Hamkins and Andy Lewis and offer an intriguing connection between computability theory and set theory. Among their special features is that the realizability model they give rise to validates the peculiar statement “there is an injection $\mathbb{R} \hookrightarrow \mathbb{N}$ ”, contradicting both classical mathematics and ordinary Turing realizability.


In the working group, we explored possible avenues for formalizing the notion of infinite time Turing machines in proof assistants based on dependent type theory. To the best of our knowledge, infinite time Turing machines have not been object of constructive scrutiny before.

In a constructive setting, a fundamental question is whether the cell contents should be Booleans or general truth values. The latter seem to force their inclusion as limsups of sequences of Booleans, but are at odds with the envisioned mechanical flavor of the computational model. We tentatively resolved this tension by sticking to Booleans, but weakening the transition function to a transition relation. Arbitrary infinite time Turing machines can then not be expected to have a well-defined execution trace, but many machines of interest will.

The working group concluded with a first sketch of the basic definitions in Agda, parametrized by the choice of ordinals used for indexing the time steps.

4.3 Tutorial on Agda, the dependently typed proof assistant

Ingo Blechschmidt (Universität Augsburg, DE)

License  Creative Commons BY 4.0 International license
© Ingo Blechschmidt

Agda is one of the commonly-used dependently typed proof assistants, facilitating collaborative verification and development of mathematical proofs. The strategic decision to schedule this tutorial on the first day of the seminar allowed participants to explore Agda during the week. The tutorial was a joint session of several people already well-versed in Agda, and profited greatly from numerous questions and comments from people with varying backgrounds, including familiarity with other different proof assistants and those new to the realm of computer formalization of mathematics.

References

- 1 M. Escardó, Introduction to Univalent Foundations of Mathematics with Agda. (2021), <https://www.cs.bham.ac.uk/~mhe/HoTT-UF-in-Agda-Lecture-Notes/>
- 2 P. Wadler, W. Kokke, J. Siek, Programming Language Foundations in Agda. (2020), <https://plfa.inf.ed.ac.uk/20.07/>

4.4 Proof systems for geometric axioms and beyond

Sara Negri (University of Genova, IT), Giulio Fellin (University of Verona, IT), Eugenio Orlandelli (University of Bologna, IT), Edi Pavlovic (LMU München, DE), and Elaine Pimentel (University College London, GB)

License  Creative Commons BY 4.0 International license
© Sara Negri, Giulio Fellin, Eugenio Orlandelli, Edi Pavlovic, and Elaine Pimentel

This working group presented an informal overview of the basics of the proof theory for geometric axioms together with recent developments.

Sara Negri presented an introduction to the method of “axioms as rules” applied to coherent and geometric axioms. By the method, cut- and contraction-free sequent calculi of the G3 family are extended while maintaining their structural properties [3, 5, 4]. One can go beyond geometric axioms by a change of basic notions (such as apartness instead of equality), by the method of systems of rules, and by the method of “geometrization” [1].

Elaine Pimentel showed how focusing and polarities can be applied for the construction of synthetic inference rules. In particular, we showed how this method can be applied for generating sequent rules for geometric axioms, advancing the work developed by Sara Negri and colleagues.

Edi Pavlović discussed the application of these methods in formal metaphysics. In particular, he discussed a four-sided sequent calculus for neutral free logic, how it simplifies quantifier rules for quantified weak Kleene, and what that can tell us.

Eugenio Orlandelli presented a terminating calculus for monadic first-order logic. The calculus has all rules invertible without relying on (implicit) contraction. The decision procedure based on this calculus has optimal complexity.

Giulio Fellin presented an infinitary version of Orevkov's theorems about the Glivenko sequent classes [2].

References

- 1 Dyckhoff, R. and Negri, S. Geometrization of first-order logic. *The Bulletin of Symbolic Logic*, vol. 21, pp. 123–163, 2015.
- 2 Fellin, G., Negri, S. and Orlandelli, E. Glivenko sequent classes and constructive cut elimination in geometric logics. *Archive for Mathematical Logic*, vol. 62, pp. 657–688, 2023.
- 3 Negri, S. Contraction-free sequent calculi for geometric theories, with an application to Barr's theorem. *Archive for Mathematical Logic*, vol. 42, pp. 389–401, 2003.
- 4 Negri, S. Geometric rules in infinitary logic. In *Arnon Avron on Semantics and Proof Theory of Non-Classical Logics, Outstanding Contributions to Logic*. Springer, pp. 265.-293, 2021.
- 5 Negri, S. and von Plato, J. *Proof Analysis: A Contribution to Hilbert's Last Problem*. Cambridge University Press, 2011.

4.5 Maps as bundles for point-free spaces

Steven J. Vickers (University of Birmingham, GB)

License  Creative Commons BY 4.0 International license
© Steven J. Vickers

Main reference Steven Vickers. Generalized point-free spaces, pointwise. arXiv:2206.01113, 2022

URL <https://arxiv.org/abs/2206.01113>

In the geometric style of point-free reasoning, just from the existence and nature of classifying toposes $\mathcal{S}[\mathbb{T}_X]$, we can say that

- *space* X = geometric theory \mathbb{T}_X (point = model of the theory),
- *map* = geometric construction of points from points.

I explained how one further gets

- *bundle* = geometric construction of spaces from points.

This is the essence of dependent type theory. We also get pullbacks of bundles as substitution. I focused on the case of localic bundles, but it does go through more generally for bounded geometric morphisms.

So: how is a map $p : Y \rightarrow X$ a bundle?

Joyal and Tierney [2] showed that localic maps p are equivalent to internal frames in the topos of sheaves over X . In [3] I showed how to replace an internal frame by a presentation \mathbb{T} of it, and then for a map $f : Z \rightarrow X$ the pullback f^*Z can be obtained from $f^*(\mathbb{T})$, where now f^* is the inverse image functor. This amounts to substitution.

If X already corresponds to a theory \mathbb{T}_X , then the internal \mathbb{T} amounts externally to an extension \mathbb{T}_Y , a process that is easier to work with if one uses “geometric type theory” rather than forcing theories into a standard form such as sites or first order theories.

For more discussion, see [1].


References

- 1 Steven Vickers. Generalized point-free spaces, pointwise. arXiv:2206.01113, 2022
- 2 Joyal and Tierney. An extension of the Galois theory of Grothendieck. *Memoirs AMS* 309, 1984
- 3 Steven Vickers. The double powerlocale and exponentiation *Theory and Applications of Categories* 12, 2004

5 Open problems

5.1 On the status of Zorn’s lemma

Ingo Blechschmidt (Universität Augsburg, DE)

License  Creative Commons BY 4.0 International license
© Ingo Blechschmidt

Over classical Zermelo-Fraenkel set theory, the axiom of choice (AC) and Zorn’s lemma (ZL) are well-known to be equivalent. Dropping the law of excluded middle (LEM) allows us to distinguish these two principles: A refined analysis shows that $AC = ZL + LEM$. While AC implies LEM and is hence a constructive taboo, ZL can be regarded as constructively neutral.

In fact, assuming ZL in the metatheory, there are plenty of models of constructive mathematics which validate ZL, and even more which validate all bounded first-order consequences of ZL: All localic Grothendieck toposes respectively all Grothendieck toposes.

That said, in mathematical practice, applications of ZL are often followed by an appeal to LEM, and without LEM, ZL loses much of its power. But there are important results which use only ZL and not LEM, such as in commutative algebra the existence of maximal ideals, the equivalence of divisible and injective abelian groups, and the existence of enough injectives.

The talk gave a summary of this circle of ideas and invited discussion on open questions: Is there a way to extract constructive content from ZL-powered results? Are there models of constructive mathematics which validate ZL, have strong ties to a given standard model and which do not require ZL in the metatheory?

5.2 Remarks on predicativity

Stefan Neuwirth (University of Franche-Comté – Besançon, FR)

License  Creative Commons BY 4.0 International license
© Stefan Neuwirth

The talk by Laura Crosilla on predicativity has triggered a special interest for the “classical approach to predicativity” that culminated in the determination of the so-called limit of predicativity. This approach has much to offer in terms of technical sophistication, but it

leads to the following question, both an epistemological and a sociological one: does the fascination for technique do justice to the very concept of predicativity? Predicativity is about the open-endedness of the process of mathematical creation; how could it be given a definite limit?

Participants

- Peter Arndt
Heinrich-Heine-Universität
Düsseldorf, DE
- Steve Awodey
Carnegie Mellon University –
Pittsburgh, US
- Andrej Bauer
University of Ljubljana, SI
- Karim Johannes Becher
University of Antwerp, BE
- Olaf Beyersdorff
Friedrich-Schiller-Universität
Jena, DE
- Marc Bezem
University of Bergen, NO
- Ingo Blechschmidt
Universität Augsburg, DE
- Ulrik Buchholtz
University of Nottingham, GB
- Gabriele Buriola
University of Verona, IT
- Felix Cherubini
Chalmers University of
Technology – Göteborg, SE
- Michel Coste
University of Rennes, FR
- Laura Crosilla
University of Florence, IT
- Nicolas Daans
Charles University – Prague, CZ
- Dominique Duval
University of Grenoble, FR
- Martín H. Escardó
University of Birmingham, GB
- Giulio Fellin
University of Verona, IT
- Makoto Fujiwara
Tokyo University of Science, JP
- Hugo Herbelin
Université Paris Cité, FR
- Matthias Hutzler
University of Gothenburg, SE
- Hajime Ishihara
Toho University – Chiba, JP
- Ulrich Kohlenbach
TU Darmstadt, DE
- Henri Lombardi
University of Franche-Comté –
Besancon, FR
- Maria Emilia Maietti
University of Padova, IT
- Julien Narboux
University of Strasbourg, FR
- Sara Negri
University of Genova, IT
- Takako Nemoto
Tohoku University – Sendai, JP
- Stefan Neuwirth
University of Franche-Comté –
Besancon, FR
- Satoru Niki
Ruhr-Universität Bochum, DE
- Paige North
Utrecht University, NL
- Eugenio Orlandelli
University of Bologna, IT
- Edi Pavlovic
LMU München, DE
- Iosif Petrakis
University of Verona, IT
- Elaine Pimentel
University College London, GB
- Michael Rathjen
University of Leeds, GB
- Marie-Françoise Roy
Université de Rennes 1 –
Rennes Cedex, FR
- Peter M. Schuster
University of Verona, IT
- Monika Seisenberger
Swansea University, GB
- Sana Stojanovic-Djurđjevic
University of Belgrade, RS
- Benno van den Berg
University of Amsterdam, NL
- Steven J. Vickers
University of Birmingham, GB



Fusing Causality, Reasoning, and Learning for Fault Management and Diagnosis

Alessandro Cimatti*¹, Ingo Pill*², and Alexander Diedrich*³

1 Bruno Kessler Foundation – Trento, IT. cimatti@fbk.eu

2 Silicon Austria Labs – Graz, AT. ingo.pill@gmail.com

3 Helmut-Schmidt-Universität – Hamburg, DE. diedrica@hsu-hh.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar “Fusing Causality, Reasoning, and Learning for Fault Management and Diagnosis” (24031). The goal of this Dagstuhl Seminar was to provide an interdisciplinary forum to discuss the fundamental principles of fault management and diagnosis, bringing together international researchers and practitioners from the fields of symbolic reasoning, machine learning, and control engineering.

Seminar January 14–19, 2024 – <https://www.dagstuhl.de/24031>

2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Theory of computation → Semantics and reasoning; Theory of computation → Theory and algorithms for application domains

Keywords and phrases cyber-physical systems, diagnosis, fault detection and management, integrative ai, model-based reasoning

Digital Object Identifier 10.4230/DagRep.14.1.25

1 Executive Summary

Alexander Diedrich (Helmut-Schmidt-Universität – Hamburg, DE)

Alessandro Cimatti (Bruno Kessler Foundation – Trento, IT)

Ingo Pill (Silicon Austria Labs – Graz, AT)

License © Creative Commons BY 4.0 International license
© Alexander Diedrich, Alessandro Cimatti, and Ingo Pill

Our goal for this Dagstuhl Seminar was to find approaches that leverage fault diagnosis to build resilient cyber-physical systems through combinations of symbolic, sub-symbolic, and control theoretic approaches.

Cyber-Physical Systems (CPSs), i.e. systems in which mechanical and electrical parts are controlled by computational algorithms, are not only continuously increasing in size and complexity, but they are also required to operate in evolving and uncertain environments, subject to frequent changes and faults. Detecting and correcting faulty behavior is a highly complex task that needs the help of computational algorithms. The constant advances in sensing technology and computational power, as well as the increase in data recording options, enables and also requires us to rely more and more on methods from Artificial Intelligence (AI) for these tasks, i.e. symbolic AI such as planning and reasoning engines, as well as subsymbolic AI like Machine Learning (ML). Sub-symbolic approaches are primarily used to detect symptoms; symbolic reasoning on the other hand provides diagnosis algorithms to identify root causes (from symptoms or observations) or reason about repairs. Furthermore,

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Fusing Causality, Reasoning, and Learning for Fault Management and Diagnosis, *Dagstuhl Reports*, Vol. 14, Issue 1, pp. 25–48

Editors: Alessandro Cimatti, Ingo Pill, and Alexander Diedrich



DAGSTUHL Dagstuhl Reports

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

control engineering methods guide the system back to normal operation (based on the identified root cause). Since these methods come from different fields, they do not always work together in practice.

The research challenge at hand is to combine symbolic a-priori knowledge and learned data, as well as to develop an integrated concept taking both symbolic and sub-symbolic approaches into account. The leading research questions of this seminar are summarised as follows:

- How can a-priori knowledge be combined with data-centric, machine learning-based algorithms?
- Can we integrate a-priori knowledge such as background knowledge about functions, interfaces and operation modes into ML-algorithms to improve model performance?
- Can we use data to learn parts of the symbolic models?
- And can we develop new algorithms which are a synthesis of both worlds, symbolic and subsymbolic?

All of these research questions must be addressed to practical and resilient cyber-physical systems. To tackle these questions, we invited researchers from symbolic AI, sub-symbolic AI, and control engineering to develop a common notion of fault detection and fault handling tasks that takes also the practical needs from industry-scale problems into account. In this regard the seminar also had a secondary function: Traditional symbolic AI diagnosis is located within the Diagnostics community (DX), while sub-symbolic fault diagnosis was traditionally associated with the fault-detection and isolation (FDI) community within the control theory research field. More recently, also the research field of machine learning has created advances with regard to fault diagnosis. Since this seminar brought together researchers from all of these fields, we hope that the seminar created fertile ground for some cross-domain research initiatives.

Besides the individual contributions to the seminar, we used four breakout sessions to brainstorm ideas and next steps following from this seminar:

1) Breakout Session on Coupling Symbolic and Sub-symbolic Methods for Model Acquisition: Fusing symbolic methods with sub-symbolic methods in both directions is essential for the creation of resilient systems. The research gap that has been identified is that so far most approaches integrate some symbolic knowledge into the majority of sub-symbolic knowledge, or a small part of sub-symbolic knowledge into a large symbolic knowledge base. But both of these directions have drawbacks and do not automatically lead to models that are well-suited for resilient systems that can be used in practice. One takeaway is the idea to organise a competition that incentivises researchers to develop novel modelling formalisms and diagnosis algorithms that mitigate some of the current drawbacks.

2) Breakout Session on Causality – How to Generate Knowledge from Data: The breakout session detailed the importance of high-level causal models in capturing causal relationships within systems. It was discussed where the difficulties in manually crafting these models lie due to their complexity and the even greater challenge of learning causal models directly from data. Crafting causal models manually, one needs a deep understanding of the dependencies. For learning causal models, a large amount of data even for situations which barely occur is needed.

3) Breakout Session on LLMs for DX – Integrating Large Language Models for Root Cause Diagnosis: The breakout session featured a comprehensive exploration and discussion on the potential and challenges of using Large Language Models in the topics of the “DX” community. The central aspects that were discussed, revolved around (i) the models themselves and their current and potential future capabilities, (ii) the training data

for training and refining LLMs for diagnosis tasks, (iii) potential application areas, as well as (iv) current, and (v) future trends and topics that should be monitored or covered by the DX community. As a result, the attendees agreed on writing a position paper, which will capture the current potential and drawbacks of LLMs within DX domains.

4) Breakout Session on Resilient Systems: For resilient systems we saw that the application and scenario play a significant role when aiming to assess what would be “good” and “bad” behavior for some system. The same goes for the question of whether we would assess the performance of a system in a local or a global context. To this end we identified a set of such relevant scenarios ranking from an energy management scenario at a local home, via the operation of an electric grid, via agents/robots in a collaborative disaster or military scenario, to supply chain management. We also discussed and converged to a definition of resilience that would tailor to all the expressed needs.

2 Table of Contents

Executive Summary

<i>Alexander Diedrich, Alessandro Cimatti, and Ingo Pill</i>	25
--	----

Overview of Talks

Keynote on Reinforcement Learning for Control of Cyber Physical Systems <i>Gautam Biswas</i>	30
Diagnosability of Fair Transition Systems <i>Marco Bozzano</i>	31
Tree-based diagnosis enhanced with meta knowledge <i>Elodie Chanthery and Louise Travé-Massuyès</i>	32
Tutorial on Runtime verification and monitor synthesis <i>Alessandro Cimatti</i>	32
AI for predictive maintenance: domain adaptation, MLOps, and Edge computing. A case study <i>Marco Cristoforetti</i>	32
Keynote on Analogy for Diagnosis by and within Cognitive Architectures <i>Kenneth D. Forbus</i>	33
Keynote on Understanding Resilience <i>Johan de Kleer</i>	34
Data-driven diagnosis from an FDI practitioner’s perspective <i>Daniel Jung</i>	34
The Rayleigh-Ritz Autoencoder architecture for Machine Learning with hard Physical Constraints <i>Manfred Mücke</i>	35
Learning what to monitor: pairing monitoring and learning <i>Angelo Montanari</i>	35
Tutorial on Diagnosing Cyber-Physical Systems <i>Oliver Niggemann</i>	36
Tutorial on Basics of Model-Based Diagnosis <i>Ingo Pill</i>	36
Designing Fault-Tolerant Control Systems using Topological Systems Theory <i>Gregory Provan</i>	37
Root Cause Analysis via Anomaly Detection and Causal Graphs <i>Josephine Rehak</i>	38
Hybrid Model Learning for System Health Monitoring <i>Pauline Ribot, and Elodie Chanthery</i>	38
Tutorial on Bridge DX / FDI <i>Louise Travé-Massuyès</i>	39
Fault Detection, Diagnosis, and Mitigation for Space Propulsion Systems <i>Günther Waxenegger-Wilfing, Kai Dresia, and Ingo Pill</i>	39

Quality Assurance Methodologies for Resilient (Model-based) Systems <i>Franz Wotawa</i>	40
Working groups	
Breakout Session on coupling symbolic and sub-symbolic methods for model acquisition <i>Rene Heesch</i>	40
Breakout Session on coupling symbolic and sub-symbolic methods in both directions <i>Rene Heesch</i>	41
Breakout Session on Causality – How to generate knowledge from data? <i>Lukas Moddemann and Kaja Balzereit</i>	42
Breakout Session on LLMs for DX – Integrating Large Language Models for Root Cause Diagnosis <i>Lukas Moddemann and Jonas Ehrhardt</i>	43
Breakout Sessions on Resilience <i>Ingo Pill</i>	44
Panel discussions	
Panel on Current and Future Challenges in Resilient System Design <i>Ingo Pill</i>	44
Open problems	
LiU-ICE Industrial Fault Diagnosis Benchmark – Anomaly Detection and Fault Isolation with Incomplete Data <i>Daniel Jung</i>	47
Participants	48

3 Overview of Talks

3.1 Keynote on Reinforcement Learning for Control of Cyber Physical Systems

Gautam Biswas (Vanderbilt University – Nashville, US)

License © Creative Commons BY 4.0 International license
© Gautam Biswas

Joint work of Gautam Biswas, Marcos Quinones-Grueiro, Austion Coursey, Avisek Naug

Main reference Avisek Naug, Marcos Quinones-Grueiro, Gautam Biswas: “Deep reinforcement learning control for non-stationary building energy management”, *Energy and Buildings*, Vol. 277, p. 112584, 2022.

URL <https://doi.org/10.1016/j.enbuild.2022.112584>

The resilience of complex cyber-physical systems (CPS) or systems of systems that combine cyber and physical components and often include humans in the loop is critical for the safe and cost-effective operations of these systems. Moreover, these systems often operate in environments whose parameters are often not known in advance, therefore, these systems have to be robust to disturbances and changes that occur in their operating environment. In other words, these systems often operate in non-stationary environments, making traditional control methods less effective for operating these systems safely and reliably.

In my presentation, I discussed the use of Reinforcement Learning (RL) methods to design controllers for complex CPS that operate in non-stationary environments. In the first third of this talk, I presented a quick introduction to RL, covering basic topics, such as Markov Decision Processes (MDPs), reward signals, value and policy functions, and the Bellman optimality criterion. I will also cover very briefly the basic RL methods of value and policy iteration, Monte Carlo and TD-learning methods, Q-learning, and Policy Gradient approaches.

In the rest of the presentation, I detailed two studies we have conducted in developing RL controllers for real-world non-stationary problems. The first is Building energy management, where we used RL to develop a supervisory controller that has been deployed on a real building for the heating, ventilation, and air conditioning (HVAC) system. Given the non-stationarities in the operating environment (e.g., sudden weather changes), we monitored for performance degradation by tracking an aggregate metric that was derived from the overall accumulated reward. Degradation in performance triggered a relearning loop. Then, a set of data-driven models of the building behavior was updated with the latest data on the building operations. Subsequently, we returned the deployed controller by letting it interact with the model and was then redeployed on the system. The approach has resulted in significant energy savings for the deployed building.

As a second case study, we developed a hybrid control framework that combines a well-established cascade control architecture and data-driven methods to accommodate varying wind conditions and payloads for unmanned aerial vehicles (UAVs). We reframed the role of the data-driven methods to compensate for the limited adaptability of the traditional control approaches by dynamically modifying the reference velocities to account for disturbances that manifest as adverse wind and payload changes. We demonstrated the advantage of the proposed framework using a Tarot T18 octocopter simulation (validated with real data) under aggressive wind field changes and payload changes mid-flight. We also showed that our learned disturbance rejection controller generalized to a different octocopter, the DJI-S1000.

The talk concluded, by discussing future work in developing continual and safe RL schemes to further enhance RL-based control and make it applicable to real-world control problems.

3.2 Diagnosability of Fair Transition Systems

Marco Bozzano (*Bruno Kessler Foundation – Trento, IT*)

License © Creative Commons BY 4.0 International license
© Marco Bozzano

Joint work of Benjamin Bittner, Marco Bozzano, Alessandro Cimatti, Marco Gario, Stefano Tonetta, Viktória Vozárová

Main reference Benjamin Bittner, Marco Bozzano, Alessandro Cimatti, Marco Gario, Stefano Tonetta, Viktória Vozárová: “Diagnosability of fair transition systems”, *Artif. Intell.*, Vol. 309, p. 103725, 2022.

URL <https://doi.org/10.1016/J.ARTINT.2022.103725>

The integrity of complex dynamic systems often relies on the ability to detect, during operation, the occurrence of faults, or, in other words, to diagnose the system. The feasibility of this task, also known as diagnosability, depends on the nature of the system dynamics, the impact of faults, and the availability of a suitable set of sensors. Standard techniques for analyzing the diagnosability problem rely on a model of the system and on proving the absence of a faulty trace that cannot be distinguished by a non-faulty one (this pair of traces is called critical pair).

In this talk, we tackled the problem of verifying diagnosability under the presence of fairness conditions. These extend the expressiveness of the system models enabling the specification of assumptions on the system behavior such as the infinite occurrence of observations and/or faults.

We adopt a comprehensive framework that encompasses fair transition systems, temporally extended fault models, delays between the occurrence of a fault and its detection, and rich operational contexts. We show that in presence of fairness the definition of diagnosability has several interesting variants, and discuss the relative strengths and the mutual relationships. We proved that the existence of critical pairs is not always sufficient to analyze diagnosability, and needs to be generalized to critical sets. We defined new notions of critical pairs, called ribbon-shape, with special looping conditions to represent the critical sets.

Based on these findings, we provide algorithms to prove the diagnosability under fairness. The approach is built on top of the classical twin plant construction, and generalizes it to cover the various forms of diagnosability and find sufficient delays.

The proposed algorithms are implemented within the xSAP platform for safety analysis, leveraging efficient symbolic model checking primitives. An experimental evaluation on a heterogeneous set of realistic benchmarks from various application domains demonstrates the effectiveness of the approach.

References

- 1 B. Bittner, M. Bozzano, A. Cimatti, M. Gario, S. Tonetta, V. Vozarova, Diagnosability of fair transition systems, *Artificial Intelligence* 309.

3.3 Tree-based diagnosis enhanced with meta knowledge

Elodie Chanthery (LAAS – Toulouse, FR) and Louise Travé-Massuyès (LAAS – Toulouse, FR)

License © Creative Commons BY 4.0 International license
© Elodie Chanthery and Louise Travé-Massuyès

Joint work of Louis Goupil, Elodie Chanthery, Louise Travé-Massuyès, Sébastien Delautier

Main reference Louis Goupil, Elodie Chanthery, Louise Travé-Massuyès, Sébastien Delautier: “Tree based diagnosis enhanced with meta knowledge”, in Proc. of the 34th International Workshop on Principles of Diagnosis (DX’23), 2023.

URL <https://hal.science/hal-04186400>

This talk presents an online data and knowledge based diagnosis method. It leverages decision trees in which decisions are made based on diagnosis meta knowledge, namely knowledge about the properties of diagnosis indicators. This knowledge is used at the level of each node to set a symbolic classification problem that brings out discriminating functions. This results in a multivariate decision tree that produces a compact model for diagnosis. The use of decision trees increases the explicability of the results found, all the more so as one discovers the explicit formal expressions of diagnosis indicators in the process. The method has been tested on static systems. On the well-known polybox, the three diagnosis indicators known as analytical redundancy relations, that are generally computed from the model, are found.

3.4 Tutorial on Runtime verification and monitor synthesis

Alessandro Cimatti (Bruno Kessler Foundation – Trento, IT)

License © Creative Commons BY 4.0 International license
© Alessandro Cimatti

Runtime Verification (RV) is a lightweight verification technique that aims at checking whether a run of a system under scrutiny (SUS) satisfies or violates a given correctness specification. The tutorial first gave an overview about the general framework of RV, and the techniques to synthesize run-time monitors that can be efficiently executed in combination with the SUS. Then, we will cover the relationship between RV and the field of Fault Detection and Isolation (FDI). In FDI, runtime monitors are built taking into account models of the SUS, in order to monitor the occurrence of internal (faulty) conditions that are not directly observable.

3.5 AI for predictive maintenance: domain adaptation, MLOps, and Edge computing. A case study

Marco Cristoforetti (Bruno Kessler Foundation – Trento, IT)

License © Creative Commons BY 4.0 International license
© Marco Cristoforetti

Joint work of Andrea Gobbi, Mario Pujatti, Diego Calzà, Piergiorgio Svaizer, Marco Cristoforetti

In recent years, data-driven artificial intelligence (AI) has acquired relevance in diagnostics. Traditional methodologies are being complemented and, in some cases, supplanted by AI-powered solutions, suggesting a possible paradigm shift. AI algorithms have demonstrated remarkable capabilities in analyzing vast amounts of data with speed and precision, enabling early detection of anomalies and predictive insights into potential faults.

The necessity of monitoring the condition of devices with diverse characteristics, each of which may exhibit unique features, behaviors, and failure modes, makes it challenging to develop a universally applicable monitoring solution based on AI that usually necessitates extensive training data. This is even more true when considering classical predictive maintenance tasks such as Remaining Useful Life (RUL) estimation, with the typical scarcity of data covering the life of the monitored system until failure. Consequently, a critical need arises for methodologies that enable these algorithms to effectively perform on unseen cases, ensuring their reliability and accuracy in practical scenarios.

Domain adaptation techniques try to solve this problem by bridging the gap between the source domain (where the model is trained) and the target domain (where it is deployed), allowing for effective knowledge transfer and adaptation to specific contexts.

This contribution presents a comprehensive, modular, and scalable solution for data-driven diagnosis and prognosis that integrates deep learning algorithms and adversarial domain adaptation to permit transfer learning and increase generalization. The pipeline starts with the data acquisition and preprocessing steps, with a configurable feature extraction phase that produces a compressed latent representation of the input samples, computed using feature extraction techniques from multiple channels of raw signals. Additional features are calculated from the latent space of deep autoencoders. This latent space is deliberately composed of only a few variables, enforcing the compression of the information contained in the input. Domain adaptation based on adversarial learning uses the features extracted from all the samples available for the source and only the first few samples from the target system. This is to align the Deep Learning regressor responsible for estimating the health index of the target necessary to compute the RUL.

In our solution, the setup includes a communication system via MQTT, enabling an online data stream for real-time monitoring and maintenance. The overall infrastructure was deployed in an industrial setting and tested in a real-time experiment, demonstrating the validity of the proposed approach.

3.6 Keynote on Analogy for Diagnosis by and within Cognitive Architectures

Kenneth D. Forbus (Northwestern University – Evanston, US)

License © Creative Commons BY 4.0 International license
© Kenneth D. Forbus

This talk described two big ideas:

1. Analogy plays key roles in human diagnosis It provides reasoning from experience and detection of novel situations Analogical generalization constructs probabilistic relational schema Provides a source of priors for model-based diagnosis Analogical learning is incremental, inspectable, and data/training efficient
2. Cognitive architectures need diagnosis
Goal: Software social organisms instead of tools Systems need to manage their own learning Achieving agency needs internal diagnostic capabilities How to build software that never blue-screens?

3.7 Keynote on Understanding Resilience

Johan de Kleer (c-infinity – Mountain View, US)

License © Creative Commons BY 4.0 International license
© Johan de Kleer

Joint work of Johan de Kleer, Alex Feldman, Ion Matei, Saigopal Nelaturi, Morad Behandesh, Ingo Pill, Jan VanderBrande, Shiwali Mohan, Wiktor Piotrowski, Sachin Grover, Sookyung Kim, Jacob Le, Roni Stern

Main reference Ion Matei, Wiktor Piotrowski, Alexandre Perez, Johan de Kleer, Jorge Tierno, Wendy Mungovan, Vance Turnewitsch: “System Resilience through Health Monitoring and Reconfiguration”, ACM Trans. Cyber Phys. Syst., Vol. 8(1), pp. 7:1–7:27, 2024.

URL <https://doi.org/10.1145/3631612>

The real world is made up of cyber-physical systems – we want them not to fail, to be invisible. How can we improve the resilience of our CPSs? Recently, a space craft designer said to me “all failures are failures of imagination.” By that he meant that it’s the designers’ responsibility to imagine all the things that could possibly go wrong with the space craft and design around or compensate for them. With the advances in AI including planning and ML we can now put AIs inside of our systems which can address the designer’s unknown unknowns. To make significant advances we need to define what resilience is and how to measure it. I will describe a number of definitions of resilience. In this talk I will describe a comprehensive approach to achieving certain types of resilience with examples ranging from printers to unmanned ships.

3.8 Data-driven diagnosis from an FDI practitioner’s perspective

Daniel Jung (Linköping University, SE)

License © Creative Commons BY 4.0 International license
© Daniel Jung

Joint work of Daniel Jung, Arman Mohammadi, Matthias Krysander

Main reference Daniel Jung: “Automated Design of Grey-Box Recurrent Neural Networks For Fault Diagnosis using Structural Models and Causal Information”, in Proc. of the Learning for Dynamics and Control Conference, L4DC 2022, 23-24 June 2022, Stanford University, Stanford, CA, USA, Proceedings of Machine Learning Research, Vol. 168, pp. 8–20, PMLR, 2022.

URL <https://proceedings.mlr.press/v168/jung22a.html>

A diagnosis system can be described as a function that uses observations from the monitored system to compute diagnoses. Because of its industrial and scientific relevance, the fault diagnosis problem has been approached in many different communities. A popular approach is data-driven fault diagnosis which refers to methods that use historical data from different fault scenarios to learn the relation between observations and diagnoses. Compared to model-based diagnosis, which uses physically based models that have a long theoretical foundation, data-driven fault diagnosis is often treated as a general classification problem. This presentation has looked at data-driven fault diagnosis from a model-based diagnosis perspective. It is shown that central model-based concepts, like redundancy, can be interpreted in a data-driven framework and it is also illustrated how these ideas can be used to develop new data-driven fault diagnosis methods.

3.9 The Rayleigh-Ritz Autoencoder architecture for Machine Learning with hard Physical Constraints

Manfred Mücke (Material Center Leoben, AT)

License © Creative Commons BY 4.0 International license
© Manfred Mücke

Joint work of Anika Terbuch, Paul O’Leary, Dimitar Ninevski, Elias Hagendorfer, Elke Schlager, Andreas Windisch, Christoph Schweimer, Matthew Harker, Manfred Mücke

Main reference Anika Terbuch, Paul O’Leary, Dimitar Ninevski, Elias Jan Hagendorfer, Elke Schlager, Andreas Windisch, Christoph Schweimer: “A Rayleigh-Ritz Autoencoder”, in Proc. of the IEEE International Instrumentation and Measurement Technology Conference, I2MTC 2023, Kuala Lumpur, Malaysia, May 22-25, 2023, pp. 1–6, IEEE, 2023.

URL <https://doi.org/10.1109/I2MTC53148.2023.10176014>

I present the Rayleigh-Ritz Autoencoder (RRAE) architecture [1] for unsupervised hybrid machine learning. It is suitable for applications where the system being observed by multiple sensors is well modeled as a boundary value problem. The embedding of the admissible functions in the decoder implements a truly physics-informed machine learning architecture. The RRAE provides an exact fulfillment of Neumann, Cauchy, Dirichlet or periodic constraints. Only the encoder needs to be trained; the RRAE is numerically more efficient during training than traditional autoencoders.

We extended the RRAE architecture [2] to distribution-free statistics to achieve stability with respect to non-Gaussian data. This provides consistent results for sensor data with both Gaussian and non-Gaussian perturbations. The necessity for handling non-Gaussian data in sensor applications is documented by the behavior of inclinometer sensors where the perturbations are characterized by Cauchy-Lorentz distribution. In such cases variance does not provide a reliable measure for uncertainty; consequently, 1-norm error measures are investigated thoroughly.

References

- 1 Anika Terbuch, Paul O’Leary, Dimitar Ninevski, Elias Hagendorfer, Elke Schlager, Andreas Windisch, and Christoph Schweimer, *A Rayleigh-Ritz Autoencoder* in 2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2023.
- 2 Anika Terbuch, Dimitar Ninevski, Paul O’Leary, Matthew Harker and Manfred Mücke. *Extended Rayleigh-Ritz Autoencoder with Distribution-Free Statistics* in 2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2024.

3.10 Learning what to monitor: pairing monitoring and learning

Angelo Montanari (University of Udine, IT)

License © Creative Commons BY 4.0 International license
© Angelo Montanari

Joint work of Angelo Montanari, Andrea Brunello, Dario Della Monica, Luca Geatti, Nicola Saccomanno

Monitoring is a runtime verification technique that can be used to check whether an execution of a system (trace) satisfies or not a given set of properties. Compared to other formal verification techniques, e.g., model checking, one needs to specify the properties to be monitored, but a complete model of the system is no longer necessary. In the talk (uploaded slides), we first introduce the notion of monitoring and display a simple architecture of a monitoring system. Then, we provide a characterisation of positively and negatively monitorable properties, and we define the safety and cosafety fragments of Linear Temporal Logic (LTL). We complete the picture by showing that monitorability goes behind safety and

cosafety LTL fragments, and that there are natural properties which are not monitorable. Next, we proceed by pointing out that monitoring suffers from some significant limitations. In particular, modern systems have such a level of complexity that it is impossible for a system engineer to specify in advance all properties to be monitored, and even minor changes to the system to be monitored can introduce unforeseen bugs. To overcome these limitations, we provide a multi-objective genetic programming algorithm to automatically extend the set of properties to monitor on the basis of the history of failure traces collected over time. The monitor and the learning algorithm are then integrated in a unifying framework, whose distinguishing features are (i) interpretability (the machine learning methods manipulate and produce only formulae, that can be easily inspected by a system engineer), (ii) formal guarantees on monitorability (every formula produced during the learning phase is guaranteed to be monitorable), and (iii) generality (different monitoring and machine learning backends). The framework has been experimentally validated on various public datasets, and the outcomes of the experimentation confirm the effectiveness of the proposed solution.

3.11 Tutorial on Diagnosing Cyber-Physical Systems

Oliver Niggemann (Helmut-Schmidt-Universität – Hamburg, DE)

License  Creative Commons BY 4.0 International license
© Oliver Niggemann

The transition from sub-symbolic representation in artificial intelligence such as time series to symbolic representations such as expressions in formal logic or in language is crucial to creating resilient cyber physical systems. The first step for this is often the discretization, i.e. the identification of symbolic concepts that come true at certain points in time. The tutorial presents typical discretization algorithms for time series, especially with a focus on engineering and scientific applications. In the last step, a brief overview of possibilities to further develop the identified concepts into causalities is given.

3.12 Tutorial on Basics of Model-Based Diagnosis

Ingo Pill (Silicon Austria Labs – Graz, AT)

License  Creative Commons BY 4.0 International license
© Ingo Pill

For Seminar 24031, we aimed to invite a diverse audience from a variety of fields connected to the seminar's focus. For us organizers it was thus important to provide the attendees with some basic knowledge, common grounds concerning well-established techniques in the field, and also a basic context and terminology. In this introductory talk, I focused on providing some brief basics about model-based diagnosis (MBD), also known as consistency-based diagnosis or DX approach. Due to its attractive features, MBD is such a central technology in the scope of this seminar, and I covered the following aspects in my presentation:

- the basic underlying approach of reasoning from first principles
- the standard scenario of explaining some unexpected behavior like a failed test case
- connections to verification tasks and techniques
- the very basic definitions of diagnoses, conflicts and minimal hitting sets
- the impact of using a weak fault model or strong fault models

- two basic algorithmic concepts for MBD: conflict-driven and direct
- MBD's flexibility in terms of application and deployed algorithm
- diagnosing multiple scenarios at the same time (e.g. like results from a test suite) and the resulting opportunity to characterize a system (when using a representative test suite, e.g., obtained with combinatorial testing)
- diagnosing static scenarios and sequential behavior
- improving the basic algorithmic concepts via algorithmic optimizations (example RC-Tree) and structural, diagnosis problem-specific information (like exploiting its parse tree when diagnosing some LTL description)
- completeness and soundness of MBD in relation to the model and the scenario(s)
- information about which authors of the covered papers participated in the seminar.

3.13 Designing Fault-Tolerant Control Systems using Topological Systems Theory

Gregory Provan (University College Cork, IE)

License © Creative Commons BY 4.0 International license
© Gregory Provan

Joint work of Gregory M. Provan, Marcos Quiñones-Grueiro, Yves Sohege

Main reference Gregory M. Provan, Marcos Quiñones-Grueiro, Yves Sohege: "Generating Minimal Controller Sets for Mixing MMAC", in Proc. of the 61st IEEE Conference on Decision and Control, CDC 2022, Cancun, Mexico, December 6-9, 2022, pp. 3009–3014, IEEE, 2022.

URL <https://doi.org/10.1109/CDC51059.2022.9993251>

Given information about (a) the desired operating conditions for a system and (b) the tasks the system must carry out, the fault-tolerant control design (FD) task is to design a set of controllers to ensure that conditions (a) and (b) are guaranteed, even when faults and/or external disturbances occur.

Designing control systems from requirements and model specifications is a challenging task, and has been addressed from many perspectives, most notably design optimization and multi-controller tuning. Our approach extends both design optimization and multi-controller tuning. Multi-controller tuning adopts a “divide and conquer” approach, decomposing the system’s operating range into smaller local sub-spaces, each associated with a “local” model. These local models are then combined to create the global system response. The primary advantage of the (MM) approach lies in its simplification of complex modeling through the use of these local models.

We develop an optimisation-based approach to designing systems with fault-tolerance and resilience capabilities, i.e., it enables multi-mode operation. Standard design optimisation approaches assume a single nominal mode; in contrast, we explicitly define an approach that generalises this framework to enable the design of systems that operate in multiple modes. Multi-controller tuning methods typically use methods to compute the set of possible modes, and then tune a controller for each mode. In contrast to multi-controller tuning methods, this new approach does not optimize just performance of individual controllers, but task-centric performance, based on topological transformations from plant spaces to control spaces.

3.14 Root Cause Analysis via Anomaly Detection and Causal Graphs

Josephine Rehak (KIT – Karlsruher Institut für Technologie, DE)

License © Creative Commons BY 4.0 International license
© Josephine Rehak

Main reference Josephine Rehak, Anouk Sommer, Maximilian Becker, Julius Pfrommer, Jürgen Beyerer: “Counterfactual Root Cause Analysis via Anomaly Detection and Causal Graphs”, in Proc. of the 21st IEEE International Conference on Industrial Informatics, INDIN 2023, Lemgo, Germany, July 18-20, 2023, pp. 1–7, IEEE, 2023.

URL <https://doi.org/10.1109/INDIN51400.2023.10218245>

In industrial processes, anomalies in the production equipment may lead to expensive failures. To avoid and avert them, the identification of the right root cause is crucial. Ideally, the search for a root cause is backed by causal information like causal graphs. We presented an extension of a framework that fuses causal graphs with anomaly detection to infer likely root causes in a process setup. The causal graph is required to contain measurement and root cause variables and causal relations with annotations for process steps. The framework uses this graph to compute for each root cause variable which measurement variable it might affect in which process step. Thereby, it considers that a cause must always precede its effect. Independently, an anomaly detection algorithm is performed on given sensor measurements to provide information about anomalies and the corresponding process step. Finally, the framework computes the likelihood of each potential root cause by comparing the results from the graph preprocessing and the anomaly detection using the Jaccard similarity to identify the most likely root cause. We demonstrated the use of this framework on a simulated robotic gripping process. Future research might investigate how to learn the causal graph from the provided data using causal discovery methods and how to apply the framework in an online fashion on given process data.

3.15 Hybrid Model Learning for System Health Monitoring

Pauline Ribot (LAAS – Toulouse, FR) and Elodie Chanthery (LAAS – Toulouse, FR)

License © Creative Commons BY 4.0 International license
© Pauline Ribot, and Elodie Chanthery

Joint work of Amaury Vignolles, Elodie Chanthery, Pauline Ribot

Main reference Amaury Vignolles, Elodie Chanthery, Pauline Ribot: “Hybrid Model Learning for System Health Monitoring”, IFAC-PapersOnLine, Vol. 55(6), pp. 7–14, 2022.

URL <https://doi.org/10.1016/j.ifacol.2022.07.098>

Health monitoring approaches are usually either model-based or data-based. This work aims at using available data to learn a hybrid model to profit from both the data-based and model-based advantages. The hybrid model is represented under the Heterogeneous Petri Net formalism. The learning method is composed of two steps: the learning of the Discrete Event System (DES) structure using a clustering algorithm (DyClee) and the learning of the continuous system dynamics using two regression algorithms (Support Vector Regression or Random Forest Regression). The method is illustrated with an academic example.

3.16 Tutorial on Bridge DX / FDI

Louise Travé-Massuyès (LAAS – Toulouse, FR)

License © Creative Commons BY 4.0 International license
© Louise Travé-Massuyès

Joint work of Marie-Odile Cordier, Philippe Dague, François Lévy, Jacky Montmain, Marcel Staroswiecki, Louise Travé-Massuyès

Main reference Marie-Odile Cordier, Philippe Dague, François Lévy, Jacky Montmain, Marcel Staroswiecki, Louise Travé-Massuyès: “Conflicts versus analytical redundancy relations: a comparative analysis of the model based diagnosis approach from the artificial intelligence and automatic control perspectives”, *IEEE Trans. Syst. Man Cybern. Part B*, Vol. 34(5), pp. 2163–2177, 2004.

URL <https://doi.org/10.1109/TSMCB.2004.835010>

Two distinct and parallel research communities have been working along the lines of the Model-Based Diagnosis approach: the FDI community and the DX community that have evolved in the fields of Automatic Control and Artificial Intelligence, respectively. This talk clarifies and links the concepts and assumptions that underlie the FDI analytical redundancy approach and the DX consistency-based logical approach. A formal framework is proposed to compare the two approaches. It is shown that by adopting the same assumptions regarding fault exoneration, they produce the same diagnostic results.

3.17 Fault Detection, Diagnosis, and Mitigation for Space Propulsion Systems

Günther Waxenegger-Wilfing (Universität Würzburg, DE), Kai Dresia (DLR – Hardthausen, DE), and Ingo Pill (Silicon Austria Labs – Graz, AT)

License © Creative Commons BY 4.0 International license
© Günther Waxenegger-Wilfing, Kai Dresia, and Ingo Pill

Joint work of Günther Waxenegger-Wilfing, Kai Dresia, Ingo Pill, Chiara Gei, Radchenko Gleb, Andrea Urgolo, Federico Pinto, Manuel Freiburger, Heike Neumann


Space propulsion systems continue to be a significant source of faults in space activities, necessitating dedicated fault management strategies to meet stringent safety requirements. The operational nature of these systems, pushed to the limits of technical feasibility to minimize weight, makes them susceptible to a diverse set of faults, with abnormal behavior having potentially catastrophic consequences. The substantial costs associated with the loss of a launch vehicle or spacecraft underscore the critical need for effective fault detection, diagnosis, and mitigation functionalities. Real-time detection and assessment of faults based on available sensor data are imperative to initiate emergency shutdowns or reconfigurations, further compounded by the constraint of limited computing resources.

The German Aerospace Center (DLR), in collaboration with various partners, has long been engaged in exploring the application of AI methods for the operation of space propulsion systems. While past efforts primarily focused on intelligent control in the absence of faults, recent initiatives, such as the collaboration with Silicon Austria Labs (SAL) within the SUNRISE project, mark the initial strides towards fault management. The SUNRISE project is dedicated to researching dependable sensor concepts for resilient control.

The first segment of this presentation unveils preliminary findings, showcasing how trained virtual sensors can effectively estimate critical quantities, such as combustion mixture ratios, and detect faults with high accuracy. In the second part, we introduce the LUMEN demonstrator engine, currently in development, serving as an ideal test bed for diverse control and diagnosis approaches. This includes intentionally injecting faults into the system and utilizing the platform for generating training data for machine learning algorithms.

3.18 Quality Assurance Methodologies for Resilient (Model-based) Systems

Franz Wotawa (TU Graz, AT)


License  Creative Commons BY 4.0 International license
© Franz Wotawa

Deploying systems requires to show that the system complies with its requirements and specifications. Hence, quality assurance is an essential part of any system development, which also holds for systems with attached resilience functionality utilizing model-based reasoning. In my talk, I discuss the general challenge of quality assurance applied to systems with monitoring and diagnosis functionality and the importance of residual risks. In particular, I mention which faults that come when introducing monitoring and diagnosis have to be considered and their effects on residual risks. Afterward, I present early work on using testing for quality assurance for models used for diagnosis and challenges. In particular, I discuss the challenge of coming up with methods for checking the quality of the tests and its consequences. Finally, I summarize the findings and present open challenges.

4 Working groups

4.1 Breakout Session on coupling symbolic and sub-symbolic methods for model acquisition

Rene Heesch (Helmut-Schmidt-Universität – Hamburg, DE)


License  Creative Commons BY 4.0 International license
© Rene Heesch

The discussion within this breakout session built upon the outcomes of the previous day, focusing on integrating prior knowledge with data to enhance model development and refinement. A critical insight from the session was the importance of creating a unified language or framework to facilitate the integration of diverse knowledge types. Additionally, it was proposed that utilizing two distinct languages or foundational models, along with a mapping between them, could also bridge the divide between symbolic and sub-symbolic methods. The session explored model checking for continuous systems and discussed algorithms for learning hybrid automata, leading to discussions on the generalizability of results from specific applications, such as ECG diagnoses, and the feasibility of learning hybrid automata directly from data. This process entails aligning a known hybrid automaton with new data and modifying the automaton based on discrepancies between the model predictions and the actual data. This method is currently being investigated by a PhD student at Laas. Furthermore, Signal Temporal Logic (STL) was discussed as a tool for applying logical systems to continuous signals or variables, demonstrating the adaptability of symbolic methods to continuous data streams. The session finally revisited the DX competition held in 2010, proposing a new competition for 2024 focusing on the development of new diagnostic models. Unlike the previous competition, which focused on algorithms, the proposed DX competition aims to challenge participants to integrate data with parts of knowledge to discover new models. Emphasis was placed on using real data from technical processes, with a focus on incorporating both nominal and faulty data into the dataset. The faulty data labels would be part of the knowledge provided to participants, allowing for a nuanced approach to model

learning and improvement. The idea is to first learn a model based on data and subsequently refine this model through the integration of additional knowledge. The session suggested not to frame this as a conventional competition but rather as a kind of special session where results could be compared and discussed, potentially leveraging past PHM (Prognostics and Health Management) challenges to minimize the preparatory work required.

4.2 Breakout Session on coupling symbolic and sub-symbolic methods in both directions


Rene Heesch (Helmut-Schmidt-Universität – Hamburg, DE)

License  Creative Commons BY 4.0 International license
© Rene Heesch

The discussion within this session was focused on exploring two main pathways for integration: a predominantly symbolic combination approach and a mainly sub-symbolic combination approach. Additionally, the concept of neuro-symbolic integration was discussed. It was clarified that sub-symbolic AI encompasses more than just Machine Learning (ML), although ML approaches were predominantly considered within the session as examples of sub-symbolic AI methods. Regarding the first pathway, the predominantly symbolic approach, which primarily utilizes sub-symbolic AI to generate models for the symbolic components, the discussion was brief. It was concluded that this topic would not be the focus due to potential overlap with another breakout session titled “Model Acquisition (Suitable for Diagnosis) from Real-World Observation”. Consequently, the discussion within in this breakout session concentrated on a primarily ML-based combination approach, outlining different key strategies. One strategy was the use of outputs from symbolic systems as inputs for ML models, providing context-rich information that is not available in raw data. This not only addresses the lack of data but also reduces reliance on large datasets by supplementing them with symbolic insights. Furthermore, the session explored incorporating logical formalisms to describe model phenomena, which enhances the interpretability and transparency of ML models in terms of their explainability. The integration of background knowledge into ML models was discussed as a crucial strategy for influencing model architecture and improving performance, particularly in data-limited scenarios or those requiring a nuanced understanding. This approach uses existing knowledge to guide the design process and boost model effectiveness. Lastly, the potential of neuro-symbolic approaches that combine neural networks with the reasoning capabilities of symbolic AI was discussed. These approaches aim to create AI systems that are not only powerful in analysis but also capable of human-like reasoning. No further steps have been defined so far, as this session merged into the session “Coupling symbolic and sub-symbolic Methods for model acquisition” later.

4.3 Breakout Session on Causality – How to generate knowledge from data?

Lukas Moddemann (Universität der Bundeswehr – Hamburg, DE) and Kaja Balzereit (Hochschule Bielefeld, DE)

License  Creative Commons BY 4.0 International license
© Lukas Moddemann and Kaja Balzereit

The breakout session on causality, held during the Dagstuhl Seminar 24031, provided a deep dive into the complexities and methodologies surrounding the identification and analysis of causal relationships within systems, especially cyber-physical systems. Several approaches to represent causality such as fault propagation graphs [Trave], causal orderings of equations [Bozzano], or causal graphs were discussed. The discussion focussed especially on fault propagation graphs, as a foundational formalism used to understand the sequential dependencies among components. These graphs are crucial for determining which components must be operational for subsequent components to function correctly, illustrating the direct causal links within a system.

A significant portion of the discussion focused on the challenges presented by cycles within fault propagation graphs. These cycles can complicate the analysis by introducing feedback loops where components influence each other in a cyclic manner, making the isolation of causal paths more complex. The session also highlighted the application of Hidden Markov Models (HMMs) as a method to model similar structures causing responses in other structures, offering a statistical approach to understanding how components influence one another even when not directly observable. The relevance of causal models for diagnosis tasks and recent work in this area [Rehak] have also been discussed.

A key takeaway from the session was the importance of high-level causal models in capturing the overarching causal relationships within systems. However, the attendees were reminded of the difficulties in manually crafting these models due to their complexity and the even greater challenge of learning causal models directly from data. When manually crafting causal models, one needs a deep understanding of the dependencies between a – usually high – number of system variables which is often not available. For learning causal models, a large amount of data even for situations which barely occur is needed. Furthermore, the distinction between correlation and causality cannot be done purely data-driven.


In conclusion, the causality breakout session provided valuable insights into the current state of causal analysis, emphasizing both the potential and the limitations of existing methodologies. The discussions underscored the need for continued research and innovation in the field to overcome the challenges of acquiring data and constructing models that can effectively capture the intricate web of causality in complex systems.

References

- 1 Trave-Massuyes, Louise, and Renaud Pons. “Causal ordering for multiple mode systems.” Proceedings of the eleventh international workshop on qualitative reasoning. 1997.
- 2 Rehak, Josephine, et al. “Counterfactual Root Cause Analysis via Anomaly Detection and Causal Graphs.” 2023 IEEE 21st International Conference on Industrial Informatics (INDIN). IEEE, 2023.
- 3 Bozzano, Marco, et al. “SMT-based validation of timed failure propagation graphs.” Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 29. No. 1. 2015.

4.4 Breakout Session on LLMs for DX – Integrating Large Language Models for Root Cause Diagnosis

Lukas Moddemann (Universität der Bundeswehr – Hamburg, DE) and Jonas Ehrhardt (Universität der Bundeswehr – Hamburg, DE)

License  Creative Commons BY 4.0 International license
© Lukas Moddemann and Jonas Ehrhardt

The breakout session on “Large Language Models for Root Cause Diagnosis” featured a comprehensive exploration and discussion on the potential and challenges of using Large Language Models in the topics of the “DX – Principles of Diagnosis” community. The central aspects that were discussed, revolved around (i) the models themselves and their current and potential future capabilities, (ii) the training data for training and refining LLMs for diagnosis tasks, (iii) potential application areas, as well as (iv) current, and (v) future trends and topics that should be monitored or covered by the DX community. As a result, the attendees agreed on writing a position paper, which will capture the current potential and drawbacks of LLMs within DX domains.

The beginning of the session included a brief introduction into the principles of LLMs. Introducing the capability of state-of-the-art LLMs and their capability on simple diagnostic benchmark problems like the Polybox. Subsequently, the prerequisites for current LLMs were discussed to effectively perform diagnoses, including necessary data, semantics, and specialized training. The capabilities of current LLMs and LLM-ensembles were discussed, highlighting the capability of formulating programming code for simple, testable and traceable reasoning, as well as the capability of understanding image data, like circuits or technical drawings, for an easier and more precise recognition of concepts of diagnosable systems. Extending the capability of pre-trained LLMs for diagnosis by fine-tuning them on diagnosis problems, as well as ensemble approaches of different Expert-LLMs for different diagnostic tasks, were highlighted as low-hanging fruits. Lastly, the capability of continuous training of LLMs for their application on changing systems was identified as a challenge.

Regarding the training data for LLMs that perform diagnostic tasks, a broad field was identified, reaching from image data, to time-series data from system observations, natural language, technical documentations, or structured knowledge graphs. The discussion came to the consensus that for training from ground up general world-knowledge should be included, whereas for fine-tuning models only specific information would be needed.

Identifying causality with LLMs was considered as a fundamental aspect of LLMs for fostering applications in diagnosis, like root cause analysis or root cause identification. Additionally, the ability to capture causality and contradiction was identified as a major aspect that should not only be considered and researched on a phenomenological level which considers LLMs as black-boxes, but also by looking into the functioning of the models.

Current trends and topics that should be monitored by the DX community revolve around the short term perspective of current LLMs in the application field of diagnosis. This includes understanding the immediate capabilities and limitation of current models, as well as application scenarios in which LLMs could already perform diagnostic tasks or at least pose as a component within a diagnosis framework.

Future trends and topics include the long term perspective of LLMs in the DX context. These trends revolve around enhancing the accuracy of LLMs and should be driven by autonomy research for structuring the requirements for LLMs in diagnostic roles.

The session concluded in writing a statement paper toward the current state of LLMs in diagnostic applications, highlighting aspects that LLMs already can achieve, and limitation they occur. The claims will be proven with empirical evaluations, like testing LLMs for creating causal graphs or evaluating on standard diagnosis problems.

4.5 Breakout Sessions on Resilience

Ingo Pill (Silicon Austria Labs – Graz, AT)

License  Creative Commons BY 4.0 International license
© Ingo Pill

In respect of these specific questions we saw that the application and scenario play a significant role when aiming to assess what would be “good” and “bad” behavior. The same goes for the question of whether we would assess the performance of a system in a local or a global context, an example for the latter being a system of resilient systems context. To this end we identified a set of such relevant scenarios ranking from an energy management scenario at a local home, via the operation of an electric grid, via agents/robots in a collaborative disaster or military scenario, to supply chain management. We also discussed and converged to a definition of resilience that would tailor to all the expressed needs. Enabled by our discussions, we identified also some follow-up actions:

- Authoring a white paper on resilience by a group of the attendees of this seminar (in 2024)
- Submitting a proposal for a follow-up Dagstuhl Seminar proposal that focuses on resilience, and where we will invite scientists from more relevant fields as well as relevant stakeholders (agency, psychology and societal sciences, security & safety, law and public regulations, ...)

Please note that the discussions led in an interdisciplinary context at Dagstuhl will also contribute to the evolution of the Int. Workshop on Principles of Diagnosis to International Conference on Principles of Diagnosis and Resilient Systems that we are implementing in 2024.

5 Panel discussions

5.1 Panel on Current and Future Challenges in Resilient System Design

Ingo Pill (Silicon Austria Labs – Graz, AT)

License  Creative Commons BY 4.0 International license
© Ingo Pill

Joint work of Ingo Pill, Gautam Biswas, Alessandro Cimatti, Johan De Kleer, Ken Forbus, Oliver Niggemann, Franz Wotawa

Directly in succession to Johan de Kleer’s keynote on Resilience discussed in an additional report, we organized this panel discussion to which we invited panelists with diverse backgrounds such as to cover topics like software engineering, intelligent agent design, automation in production and manufacturing, AI-based control, rigorous system design, formal verification, run-time verification and monitoring, intelligent sensing, and other related topics that are related to the diverse challenges connected to designing resilient systems. Similar to the term “artificial intelligence”, there seems to be an intuitive understanding of the concept’s purpose and the meaning of resilience on an abstract level. As we saw in our discussions, there are, however, also differences in how to interpret the concept and what to expect from a resilient system. As an initial characterization, let us thus describe resilience as a system’s intrinsic ability of sustaining its operation also when impacted by anticipated and unexpected contingencies. In this context, we would like to distinguish between basic and extreme resilience as follows: Basic resilience would allow a system to cope with anticipated

issues, while extreme resilience would enable a system to deal also with challenges that were not anticipated when the system was designed. While resilience could relate also to resilient design concepts that would allow a designer or developer to react more easily to design/requirement changes (or that certain components are resilient to changes in other components). In contrast, we consider the major focus of resilience to be on maintaining expected operation during its operation, no matter the circumstances. It is important to note though that we have to design a system such as to add the capabilities of dealing with (unexpected) issues (faults, threats, environmental changes, ...) at design time – it is only the effects achieved that we are experiencing at run-time. To the end of discussing relevant technologies, we invited our panelists to give short lightning talks where we tasked them to provide some background information about resilience aspects in their individual expertise to the audience, and to comment on the most recent questions and thoughts covered by frontier research. Including the discussion among the panelists, with the moderator and also the entire audience, the lightning talks inspired the following discussions:

- Gautam Biswas brought up in his statement the fact that designing resilience into a system can be thought about in two directions. That is, we can anticipate issues and design a system in a way that it would be “robust enough” against certain problems by design. The second concept would be to allow a system to assess and consider a situation at run-time, reason about an appropriate mitigation strategy and then take mitigation actions – all done at run-time. We can emphasize on the second option when using the term operational resilience. There are several stages that a system goes through in the context of such operational resilience, in that a system would suffer from degraded performance before the mitigation strategy’s effects manifest and the system’s desired performance is restored (to the degree to which this is possible).
- Alessandro Cimatti focused in his statement on the challenge of defining resilience, and he referred to multiple example domains for showcasing relevant questions. He brought up the implicit connection to fail-operational concepts, to the operation of adaptive systems, to planning in the context of non-determinism and uncertain duration, and he observed that such planning alone won’t go far on its own. That is, it is a combination of techniques that will be necessary to tackle the challenges faces when aiming for resilience in a system’s behavior (like the ability to extract models for evolving dynamic environments). A specific question of interest is that of enabling resilience from a short- and a long-term perspective
- Ken Forbus focused in his statement titled “Analogy and Cognitive Architecture as Sources of Software Resilience” on the importance of the concept of analogy, as well as the design of a cognitive architecture propelling the performance in resilient, intelligent systems. Especially in the context of extreme resilience, a system faces incomplete information (not only in terms of the environment, but also referring to domain knowledge) like when we initially did not know how to deal with Covid-19 as a society. Drawing on analogies and exploiting earlier expertise for analogous situations could be one fundamental technology to drive solutions for achieving resilience. This will require us to enable agency in a system, such that systems will elevate from being simple tools to becoming intelligent and evolving agents. The cognitive architectures that would allow us in implementing such agency (that then enables a system to efficiently come up and effectively execute appropriate mitigation strategies) are among the currently most relevant challenges.
- Oliver Niggemann discussed the necessity of considering backloading instead of frontloading when designing a resilient system, and that we need to adapt our design processes as well as the education of engineers accordingly. In particular we see that, while a well-engineered system is a prerequisite for a system to be trustworthy, the complexity and

dynamics of applications requires us to come up with trustworthy AI-based solutions for operating a system. Designers and engineers thus will need to think about the operation phase in more detail when developing future systems. Our education and engineering concepts have to be adapted in order to support and being able to leverage resilient system design in practical designs. This includes also addressing the fact that we have currently a set of technologies available that are promising in being able to address one or the other resilience aspect from a scientific perspective. We lack, however, integrated approaches and methodologies that we can then deploy in practice and transfer to an industrial context, so that a big challenge in this context of resilience is to develop those.

- Franz Wotawa discussed in his statement three fundamental questions, considering resilience not only from a design perspective but also from the perspective of evaluating a resilient system design: What is the best resilient system design? What are desired properties of resilient systems? How do we ensure the correctness of resilient systems? Addressing those questions requires us to think about architectural aspects but also about our development processes that now have to facilitate resilience in a design. This entails the need to establish not only a common understanding of the purpose and meaning of resilience, but also of the degrees of freedom a resilient system is allowed to operate within. This is crucial not only from a design perspective, but especially so in an evaluation and verification context. Enabling the latter, we need to come up with concrete evaluation metrics, and we need to define exact bounds of acceptable autonomy in resilience.


From the discussions we had in the panel and two break-out sessions, we can immediately conclude that achieving resilience is a very complex task. Aside apparent technical and technological questions, there are also legal ones, like who would be responsible if the required autonomy to achieve operational resilience causes harm, damage, or the loss of revenue. Psychological and societal questions are also relevant, like in the sense that they influence the definition of acceptable behavior or are of interest in a technological context when we think of human-machine collaboration or also systems of resilient systems where we could have humans in the loop or where we are (at the very least) operating in a shared environment. While some resilience can be achieved with anticipating challenges and making a design resilient by default to those, it is especially the extreme resilience where we equip a system with the intelligence to overcome unexpected issues at run-time that requires us to rethink our current design, development, evaluation and verification processes. Due to the complexity of the discussions, they were led not only in the time-limited context of the panel discussion, but specific aspects were discussed also in two break-out sessions:

- What is resilience and how do we measure it?
- Dealing with unknown unknowns in resilience?

6 Open problems

6.1 LiU-ICE Industrial Fault Diagnosis Benchmark – Anomaly Detection and Fault Isolation with Incomplete Data

Daniel Jung (Linköping University, SE)

License  Creative Commons BY 4.0 International license

© Daniel Jung

Joint work of Daniel Jung, Mattias Krylander, Erik Frisk

URL https://vehsys.gitlab-pages.liu.se/diagnostic_competition/

A common challenge of designing diagnosis systems in industrial applications, is limited data availability from relevant fault scenarios and a lack of knowledge of model uncertainty. Development of fault diagnosis design techniques in this situation is the theme of the competition.

The case study is the air-flow of an internal combustion engine. The complexity of modeling the engine together with noisy measurements makes is a challenging system to diagnose because of its non-linear dynamic behavior and wide operating range.

Competition Objectives

- Design a diagnosis system that can detect and isolate faults.
- Handle that availability of representative data from all fault scenarios and fault sizes is limited.
- The diagnosis system should handle faults that are not represented in training data.

Participants

- Kaja Balzerit
Hochschule Bielefeld, DE
- Gautam Biswas
Vanderbilt University –
Nashville, US
- Marco Bozzano
Bruno Kessler Foundation –
Trento, IT
- Elodie Chanthery
LAAS – Toulouse, FR
- Alessandro Cimatti
Bruno Kessler Foundation –
Trento, IT
- Marco Cristoforetti
Bruno Kessler Foundation –
Trento, IT
- Philippe Dague
University Paris-Saclay –
Orsay, FR
- Johan de Kleer
c-infinity – Mountain View, US
- Alexander Diedrich
Helmut-Schmidt-Universität –
Hamburg, DE
- Kai Dresia
DLR – Hardthausen, DE
- Jonas Ehrhardt
Universität der Bundeswehr –
Hamburg, DE
- Alexander Feldman
NextFlex – San Jose, US
- Kenneth D. Forbus
Northwestern University –
Evanston, US
- Rene Heesch
Helmut-Schmidt-Universität –
Hamburg, DE
- Daniel Jung
Linköping University, SE
- Lukas Moddemann
Universität der Bundeswehr –
Hamburg, DE
- Angelo Montanari
University of Udine, IT
- Manfred Mücke
Material Center Leoben, AT
- Edi Muskardin
Silicon Austria Labs – Graz, AT
- Oliver Niggemann
Helmut-Schmidt-Universität –
Hamburg, DE
- Ingo Pill
Silicon Austria Labs – Graz, AT
- Gregory Provan
University College Cork, IE
- Belarmino Pulido
University of Valladolid, ES
- Josephine Rehak
KIT – Karlsruher Institut für
Technologie, DE
- Pauline Ribot
LAAS – Toulouse, FR
- Martin Sachenbacher
Universität Lübeck, DE
- Anika Schumann
IBM Research-Zurich, CH
- Gerald Steinbauer-Wagner
TU Graz, AT
- Markus Stumptner
University of South Australia –
Mawson Lakes, AU
- Anna Szyber
Warsaw University of
Technology, PL
- Louise Travé-Massuyès
LAAS – Toulouse, FR
- Günther Waxenegger-Wilfing
Universität Würzburg, DE
- Katinka Wolter
FU Berlin, DE
- Franz Wotawa
TU Graz, AT
- Marina Zanella
University of Brescia, IT
- Alois Zoitl
Johannes Kepler Universität
Linz, AT



Report from Dagstuhl Seminar 24032

Representation, Provenance, and Explanations in Database Theory and Logic

Pablo Barcelo^{*1}, Pierre Bourhis^{*2}, Stefan Mengel^{*3}, and Sudeepa Roy^{*4}

- 1 PUC – Santiago de Chile, CL. pbarcelo@ing.puc.cl
- 2 CNRS – CRIStAL, Lille, FR. pierre.bourhis@univ-lille.fr
- 3 CNRS, CRIL – Lens, FR. mengel@cril.fr
- 4 Duke University – Durham, US. sudeepa@cs.duke.edu

Abstract

This report documents the program and the outcomes of **Dagstuhl Seminar “Representation, Provenance, and Explanations in Database Theory and Logic” (24032)**, which was broadly in the area of database theory. Database theory formalizes the theoretical underpinnings of databases and analyzes them with mathematical tools. We focused on questions related to the fundamental problem of efficient query evaluation: compute the answers of a query on a database. This seminar focused on three key aspects of query evaluations. (1) **Representation** studies the tradeoff between expressivity, compactness, and efficient computation of outputs from the inputs, including circuits and knowledge compilation forms, enumeration, and direct access. (2) **Provenance** captures the computation process of outputs from the inputs using a compact formula, and has applications to probabilistic databases. (3) **Explanations** give meaningful insights to responsibilities of different inputs toward an output beyond provenance, e.g., by using Shapley Values from co-operative game theory that has been recently popular in both DB and ML.

Seminar January 14–19, 2024 – <https://www.dagstuhl.de/24032>

2012 ACM Subject Classification Theory of computation → Theory and algorithms for application domains; Theory of computation → Theory and algorithms for application domains; Theory of computation → Theory and algorithms for application domains; Theory of computation → Theory and algorithms for application domains; Theory of computation → Theory and algorithms for application domains

Keywords and phrases Circuits, database theory, factorized databases, provenance, shapley values

Digital Object Identifier 10.4230/DagRep.14.1.49

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Representation, Provenance, and Explanations in Database Theory and Logic, *Dagstuhl Reports*, Vol. 14, Issue 1, pp. 49–71

Editors: Pablo Barcelo, Pierre Bourhis, Stefan Mengel, and Sudeepa Roy



DAGSTUHL Dagstuhl Reports

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


1 Executive Summary

Pablo Barcelo (PUC – Santiago de Chile, CL)

Pierre Bourhis (CNRS – CRISTAL, Lille, FR)

Stefan Mengel (CNRS, CRIL – Lens, FR)

Sudeepa Roy (Duke University – Durham, US)

License  Creative Commons BY 4.0 International license
 © Pablo Barcelo, Pierre Bourhis, Stefan Mengel, and Sudeepa Roy

Background and Research area

The Dagstuhl Seminar “*Representation, Provenance, and Explanations in Database Theory and Logic*” (24032) was broadly in *Database Theory*, where the goal is to formalize the theoretical underpinnings of databases and then analyze them with mathematical tools. One of the most fundamental problems in both database theory and systems is efficient query evaluation: given a database and a query, compute the answer to the query on the database. This question has a tight connection to logic, since it has been known for a long time that different fragments of first- or second-order logic can be seen as the core of practical query languages like SQL or Datalog. This seminar focused on three key aspects of query evaluations: *representation*, *provenance*, and *explanations*.

Representation. For large datasets, query results can be very large when they are materialized explicitly in the standard form. For efficient query processing and subsequent applications, it is important to *represent* the query answers in a compact fashion. One important form of representations in query evaluation is by *circuits*, which have a long history in complexity theory and AI and can be seen as part of the larger framework of *knowledge compilation* (Darwiche, Marquis, J. Artif. Intell. Res. 2002). Circuits were heavily discussed in several presentations in the seminar. The other aspect of representation that the seminar focused on was the field of *enumeration algorithms* and *direct access*. It first computes a data structure representing the query answers, and then gives an algorithm to extract one answer at a time from the data structure. In this problem, the complexity of the two parts is measured separately: the computation time of the data structure is called the *preprocessing time* and the time of the extraction of each answer is called the *delay*. Typically, the goal of such algorithms is to have a preprocessing time much smaller than the cost of the classical evaluation of the query and very small (ideally constant) delay.

Provenance. Data provenance in general refers to how the outputs of a query are generated from the inputs, with a broad goal to enable interpretability, trust, and reproducibility of the queries. A mathematical form of provenance that propagates annotations of inputs to the outputs, called *provenance semirings*, was proposed in a seminal work by Green et al. (PODS 2007). The most specialized case of Boolean semirings captures how an output tuple has been obtained from the inputs with joint usage (joins – translate to conjunctions \wedge), and alternative usages (projections or unions – translate to disjunctions \vee). Such semirings can be used to understand compactly how outputs are generated from inputs, and have applications in *query evaluation in probabilistic databases* when realization of inputs tuples is uncertain (Dalvi-Suciu, JACM 2012), and in *deletion propagation* or *view update*, to understand how the outputs change if one or more inputs are deleted, without re-computing the query. There are more advanced semirings like tropical semirings that can capture shortest paths in graphs. Compact and efficient knowledge compilations of provenance circuits into *ordered and free binary decision diagrams (OBDDs, FBDDs)*, and more generally as *decomposable*

deterministic negation normal forms (d-DNNF) are also important questions in database theory with applications in probabilistic databases (Jha-Suciu, ICDT 2011; Beame et al., ACM Trans. Database Syst. 2017; Monet, PODS 2020).

Explanations. While provenance provides one approach to explaining query answers capturing how the query answers are generated, in many applications, other forms of insights as explanations are desired for understanding contributions of inputs, trends and anomalies in the outputs, and deciding next course of actions or recourse. Recently, explanations based on the widely known *Shapley values* from co-operative game theory have been used in database theory to measure the relevance of a certain database fact to a query answer (Deutch et al, SIGMOD 2022; Livshits et al., ICDT 2021), and to measure the relevance of inputs to the outcome of an ML classifier (Arenas et al., AAAI 2021). Complexity, applications, and algorithms for explanations by Shapley values were heavily discussed in the seminar. Since the naive computation of Shapley values is intractable as it includes a summation over exponentially many subsets, one of the main themes behind this investigation has been the identification of practically relevant classes of database queries for which such explanations can be computed in polynomial time, possibly using knowledge compilation forms. Apart from Shapley values, other forms of explanations, including that of aggregated database query answers (e.g., Roy-Suciu, SIGMOD'14) and connections of explanations with data privacy, fairness, and causal inference were discussed in the seminar. This way the seminar connected the field of database theory to the field of *responsible data science* that is of paramount importance in real world.

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.

2 Table of Contents

Executive Summary	
<i>Pablo Barcelo, Pierre Bourhis, Stefan Mengel, and Sudeepa Roy</i>	50
Organization of the Seminar	54
Outcomes of the seminar	54
Overview of Talks	56
Consistency of Relations over Monoids	
<i>Albert Atserias</i>	56
Lower Bounds on Probabilistic Query Evaluation	
<i>Antoine Amarilli</i>	56
Privately Generating Justifiably Fair Data	
<i>Amir Gilad</i>	57
Model Interpretability through the Lens of Computational Complexity	
<i>Pablo Barcelo</i>	57
SHAP-Scores and Its Computation over ML Models	
<i>Pablo Barcelo</i>	58
A dichotomy for succinct representations of homomorphisms	
<i>Christoph Berkholz</i>	58
Tractability and Optimization of Shap-Score Computation for Explainable AI	
<i>Leopoldo Bertossi</i>	59
Circuits for Query Evaluation over Trees	
<i>Pierre Bourhis</i>	59
From Queries to Circuits	
<i>Florent Capelli</i>	60
Unified Reverse Data Management	
<i>Wolfgang Gatterbauer</i>	60
Graph Explainability and Shapley	
<i>Floris Geerts</i>	61
Lessons Learned from Building Systems for Provenance and Explanations	
<i>Boris Glavic</i>	61
Answering Database Queries Using Direct-Access Structures	
<i>Benny Kimelfeld</i>	62
The Shapley Value in Database Management	
<i>Ester Livshits</i>	62
Provenance in Queries, Games, and Argumentation: Time for a Family Reunion	
<i>Bertram Ludäscher</i>	63
Impact of Self-Joins on Enumeration and Direct Access on Join Queries	
<i>Stefan Mengel</i>	64
The Intensional-Extensional Problem in Probabilistic Databases	
<i>Mikaël Monet</i>	64

Revisiting Semiring Provenance for Datalog <i>Liat Peterfreund</i>	65
Datalog over (Pre-)Semirings <i>Reinhard Pichler</i>	65
MSO Enumeration over Words and their Representations <i>Cristian Riveros</i>	66
Explanations for Aggregate Query Answers – An Overview <i>Sudeepa Roy</i>	67
Training Invariant Machine Learning Models with Incomplete Data <i>Babak Salimi</i>	67
Expected Shapley-Like Scores of Boolean Functions: Complexity and Applications to Probabilistic Databases <i>Pierre Senellart</i>	68
The Importance of Parameters in Database Queries <i>Christoph Standke</i>	68
From Shapley Value to Model Counting and Back <i>Dan Suciu</i>	69
Answering Quantile Join Queries by Representing Inequality Predicates Efficiently <i>Nikolaos Tziavelis</i>	69
Open problems	70
A problem on unambiguous DNFs <i>Mikaël Monet</i>	70
A question about representability of probabilistic databases <i>Christoph Standke</i>	70
Participants	71

3 Organization of the Seminar

The seminar was held between January 14–19, 2024 (Monday to Friday with arrival on Sunday). We had 26 on-site participants. We started the first day with an introduction of each participant presenting their background, research area, as well as what they wished to achieve from the seminar. Right after, we had the opening keynote (the only one-hour talk) of the seminar by Reinhard Pichler on Datalog over semirings. We had a mix of 45 mins, 30 mins, and 20 mins talks in the rest of the seminar. On the first day, we had 8 talks of different length, focusing on backgrounds on semirings, explanations for database queries and explanations in ML, and Shapley values, and short talks on various topics later in the day. The aim was to cover a significant part of the background for the rest of the seminar as well as to learn about interesting research from several of the participants on the very first day. This allowed us to have more relaxed schedule in the rest of the week with more time for free collaboration, as well as to schedule more technical talks later in the week. On Tuesday and Wednesday, we had talks on model counting, probabilistic databases, enumeration, direct access, semirings, and circuits. On Thursday, we focused on systems and application aspects, including causal inference, fairness, and privacy, and short talks on miscellaneous topics. We had ample time of free discussions from Tuesday to Friday (including the typical time for excursion on Wednesday afternoon, which had to be canceled because of bad weather). We saw talks ranging from logic and complexity, systems, to applications related to the seminar topics. There were 26 talks spread over the first four days of the seminar given by 25 participants, and two open problem sessions (Thursday morning and Friday morning). All talks were well received, with many questions and lively discussions during and after the talks. Overall, the seminar was highly engaging, intellectually stimulating, and a great success.

4 Outcomes of the seminar

- **Scientific content:** The participants learnt about backgrounds and recent work on the seminar topics – *representations, provenance, and explanations*, from experts. In the opening keynote, Reinhard Pichler gave a comprehensive introduction to *provenance semirings* that was used in a large number of talks in the seminar. He also talked about their recent work on the query language Datalogo, which is based on the concept of K-relations and generalizes recursive Datalog to (pre-)semirings. Later, we saw talks on different semantics of provenance semirings for Datalog (Liat Peterfreund), consistency of relations over monoids (Albert Atserias), and provenance in queries, games, and argumentation (Bertram Ludascher).

We saw different views and applications of *explanations* in the seminar. Sudeepa Roy talked about explanations for aggregate query answers: in response to user questions on why an output is high/low, or higher/lower than other attributes, how to generate deep explanations that the domain experts can find from the data automatically. Pablo Barcelo talked about another form of explanations, a framework for judging and comparing the interpretability of different ML models, and complexity of this problem. A number of presentations discussed various aspects of *Shapley (SHAP) values* as explanations: computing SHAP scores over ML models (Pablo Barcelo), using Shapley values to measure the responsibility of individual database tuples to the outcome for query answering and database inconsistency (Ester Livshits), tractability of SHAP scores for explainable AI (Leo Bertossi), use of Shapley-like scores for explaining graph neural networks (Floris

Geerts), polynomial-time equivalence between computation of Shapley values and model counting for a class of functions (Dan Suciu), equivalent tractability of the computations of expected Shapley values and of the expected values of Boolean functions in probabilistic databases (Pierre Senellart), and quantifying the importance of the choices of parameter values to the result of a query over a database using SHAP scores (Christoph Standke). We had several talks on *query evaluations on probabilistic (uncertain) data*, which made connections between provenance polynomials and their representations as circuits. Antoine Amarilli talked about lower bounds for probabilistic query evaluation, for both computation and size of provenance as circuits. Mikael Monet revisited the intensional-extensional problem in probabilistic databases, and talked about their ongoing work on whether the tractability for UCQ can be captured by knowledge compilation. Pierre Bourhis talked about circuits for query evaluation over trees, and Florent Capelli discussed algorithms to construct tractable circuits from queries.

For *representations* we had multiple talks on direct access and enumeration. Stefan Mengel talked about the impact of self-join for such queries. Cristian Riveros presented a survey of MSO enumeration problems over words based on the model of annotated automata. Benny Kimelfeld discussed fine-grained complexity of database queries that involve joins, grouping, aggregation, and ordering using direct access structures. Nikos Tziavelis talked about the complexity of answering quantile join queries by efficiently representing inequality predicates.

Wolfgang Gatterbauer made a connection between the problem of finding minimal size *provenance factorizations* and reverse data management problems such as resilience (how to change a query answer with smallest change in data). Christoph Berkholz discussed lower bounds for factorized representations for multi-way join queries and homomorphisms between two structures.

On the *systems and applications* side, Boris Glavic shared with the participants the lessons he learned from his work on building systems for capturing and managing provenance and explanations, and the separation between data flow between operators in a query and the information flow by provenance. Making connections with responsible data science, Babak Salimi discussed the challenges and solutions in training ML models with incomplete data and in the presence of selection bias in data, and its applications in the context of fairness. Amir Gilad presented a framework for synthetic data generation that is both differentially-private and fair.

- **Open problems:** The participants discussed several open problems in the open problem sessions. For instance, (1) what are the notions equivalent to semirings for datalog with negation? Earlier, monus has been proposed for non-monotone queries. How do the requirements for being stable (from the recent work on Datalogo over semirings) and having a monus interact? (2) How do we define and complexity for Shapley values for queries with negation and queries with aggregates? Do the approximation results from the recent literature still hold when we have negation? For queries with aggregates, should we assign responsibilities to single tuples or a group of tuples? Two other open problems are listed at the end of this document.
- **Making connections between seminar topics and theory, systems, and applications:** The seminar brought together researchers who are broadly interested in one or more of the seminar topics, but work on different aspects of these topics. While a majority of the participants work on the theoretical aspects of the topics, some participants work on systems and the other work on applications and data science. We also saw interesting exchanges of ideas among different topics (representations, provenance, and explanations) in the seminar.

- **Extensive collaborations:** In addition to learning about recent research from the talks, the participants extensively discussed problems with old or new collaborators during the week.

5 Overview of Talks

5.1 Consistency of Relations over Monoids

Albert Atserias (UPC Barcelona Tech, ES)

License © Creative Commons BY 4.0 International license
© Albert Atserias

Joint work of Albert Atserias, Phokion G. Kolaitis

Main reference Albert Atserias, Phokion G. Kolaitis: “Consistency of Relations over Monoids”, CoRR, Vol. abs/2312.02023, 2023.

URL <https://doi.org/10.48550/ARXIV.2312.02023>

The interplay between local consistency and global consistency has been the object of study in several different areas, including probability theory, relational databases, and quantum information. For relational databases, Beeri, Fagin, Maier, and Yannakakis showed that a database schema is acyclic if and only if it has the local-to-global consistency property for relations, which means that every collection of pairwise consistent relations over the schema is globally consistent. More recently, the same result has been shown under bag semantics. In this paper, we carry out a systematic study of local vs. global consistency for relations over positive commutative monoids, which is a common generalization of ordinary relations and bags. Let K be an arbitrary positive commutative monoid. We begin by showing that acyclicity of the schema is a necessary condition for the local-to-global consistency property for K -relations to hold. Unlike the case of ordinary relations and bags, however, we show that acyclicity is not always sufficient. After this, we characterize the positive commutative monoids for which acyclicity is both necessary and sufficient for the local-to-global consistency property to hold; this characterization involves a combinatorial property of monoids, which we call the *transportation property*. We then identify several different classes of monoids that possess the transportation property. As our final contribution, we introduce a modified notion of local consistency of K -relations, which we call *pairwise consistency up to the free cover*. We prove that, for all positive commutative monoids K , even those without the transportation property, acyclicity is both necessary and sufficient for every family of K -relations that is pairwise consistent up to the free cover to be globally consistent.

5.2 Lower Bounds on Probabilistic Query Evaluation

Antoine Amarilli (Telecom Paris, FR)

License © Creative Commons BY 4.0 International license
© Antoine Amarilli

This talk focuses on the task of computing the probability that a fixed query holds on an input probabilistic database. The problem can also be specialized to several contexts, e.g., computing the probability that an input graph with probabilistic edges contains a specific pattern, or in the unweighted case counting how many subgraphs of the input have a certain property. We will review recent hardness results on this problem. We will cover two kinds of

results: lower bounds on the computational complexity of the problem, and lower bounds on the size of the query provenance when represented in structured circuit classes.

References

- 1 Antoine Amarilli, Timothy van Bremen, Kuldeep S. Meel: Conjunctive Queries on Probabilistic Graphs: The Limits of Approximability. ICDT 2024.
- 2 Antoine Amarilli. Uniform Reliability for Unbounded Homomorphism-Closed Graph Queries. ICDT 2023.
- 3 Antoine Amarilli, Benny Kimelfeld. Uniform Reliability of Self-Join-Free Conjunctive Queries. LMCS, 2022.
- 4 Antoine Amarilli, Mikaël Monet. Weighted Counting of Matchings in Unbounded-Treewidth Graph Families. MFCS 2022.

5.3 Privately Generating Justifiably Fair Data

Amir Gilad (The Hebrew University of Jerusalem, IL)

License © Creative Commons BY 4.0 International license
© Amir Gilad

Joint work of David Pujol, Amir Gilad, Ashwin Machanavajjhala

Main reference David Pujol, Amir Gilad, Ashwin Machanavajjhala: “PreFair: Privately Generating Justifiably Fair Synthetic Data”, Proc. VLDB Endow., Vol. 16(6), pp. 1573–1586, 2023.

URL <https://doi.org/10.14778/3583140.3583168>

In this talk, I will present our recent work that develops a framework for synthetic data generation that is both differentially-private and fair, where fairness is modeled by an adaptation of the causal definition for justifiable fairness.

5.4 Model Interpretability through the Lens of Computational Complexity

Pablo Barcelo (PUC – Santiago de Chile, CL)

License © Creative Commons BY 4.0 International license
© Pablo Barcelo

Joint work of Pablo Barcelo, Bernardo Subercaseaux

This talk revisits a framework for judging and comparing the interpretability of classes of Machine Learning models. Said framework allows us to formalize and prove a nuanced version of claims like “decision trees are more interpretable than neural networks”. Interestingly, such a formalization pointed out the first result establishing the hardness of interpreting decision trees, and provided tools to analyze how hyper-parameters such as the number of layers in a network can impact its interpretability. Our framework relied on a few assumptions that will be discussed explicitly in the talk, such as the role of well-defined interpretability queries or the adequacy of computational complexity for capturing the practical complexity of real-life instances.

References

- 1 Model Interpretability through the Lens of Computational Complexity. Pablo Barceló, Mikaël Monet, Jorge Pérez, Bernardo Subercaseaux. (<https://arxiv.org/abs/2010.12265>)
- 2 On Computing Probabilistic Explanations for Decision Trees. Marcelo Arenas, Pablo Barceló, Miguel Romero, Bernardo Subercaseaux. (<https://arxiv.org/abs/2207.12213>)

5.5 SHAP-Scores and Its Computation over ML Models

Pablo Barcelo (PUC – Santiago de Chile, CL)

License  Creative Commons BY 4.0 International license
 © Pablo Barcelo

Main reference Marcelo Arenas, Pablo Barceló, Leopoldo E. Bertossi, Mikaël Monet: “On the Complexity of SHAP-Score-Based Explanations: Tractability via Knowledge Compilation and Non-Approximability Results”, *J. Mach. Learn. Res.*, Vol. 24, pp. 63:1–63:58, 2023.

URL <http://jmlr.org/papers/v24/21-0389.html>

SHAP scores are expressions designed to capture the contribution of a feature to the output of a machine learning model. They are grounded in the well-studied game-theoretical notion of Shapley values. In this discussion, I will elucidate the meaning of these SHAP scores expressions and explain how they are obtained from first principles. Subsequently, I will delve into the examination of the problem of computing SHAP scores over machine learning models. I will provide insights into when and why this problem becomes computationally intractable. Additionally, I will identify a large and practically relevant class of models for which the problem can be solved in polynomial time. Finally, I will show that even for slight extensions of this class, the computation of SHAP scores is not only intractable but also does not admit a Fully Polynomial Randomized Approximation Scheme (FPRAS).

5.6 A dichotomy for succinct representations of homomorphisms

Christoph Berkholz (TU Ilmenau, DE)

License  Creative Commons BY 4.0 International license
 © Christoph Berkholz

Main reference Christoph Berkholz, Harry Vinnal-Smeeth: “A Dichotomy for Succinct Representations of Homomorphisms”, in *Proc. of the 50th International Colloquium on Automata, Languages, and Programming, ICALP 2023, July 10-14, 2023, Paderborn, Germany, LIPIcs, Vol. 261*, pp. 113:1–113:19, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023.

URL <https://doi.org/10.4230/LIPICS.ICALP.2023.113>

The talk is based on the cited ICALP’23 paper. It will be about factorized databases for multi-way join queries, or, in other words, succinct representations of all homomorphisms between two structures A and B. The main result is a characterisation of (bounded-arity) structures A where this is efficiently doable. In the talk I will mainly focus on lower bounds for factorized representations.

5.7 Tractability and Optimization of Shap-Score Computation for Explainable AI

Leopoldo Bertossi (SKEMA Business School – Montréal, CA)

License © Creative Commons BY 4.0 International license
© Leopoldo Bertossi

Joint work of Marcelo Arenas, Pablo Barceló, Mikäel Monet, Jorge León

Main reference Leopoldo E. Bertossi, Jorge E. León: “Efficient Computation of Shap Explanation Scores for Neural Network Classifiers via Knowledge Compilation”, in Proc. of the Logics in Artificial Intelligence – 18th European Conference, JELIA 2023, Dresden, Germany, September 20-22, 2023, Proceedings, Lecture Notes in Computer Science, Vol. 14281, pp. 49–64, Springer, 2023.

URL https://doi.org/10.1007/978-3-031-43619-2_4

The presentation is about recent research on the Shap Scores in Explainable Machine Learning. More specifically, on the basis of the tractability result for Shap [1] for open-box classifiers defined by a class of Boolean circuits (actually, d-DBC), we show how Shap can be computed much more efficiently than through the sheer use of the classifier’s input/output relation when a Binary Neural Network classifier is, first, represented by means of a compact CNF formula, which is, next, (knowledge) compiled into an SDD, followed by a transformation into a d-DBC [2].

References

- 1 Marcelo Arenas, Pablo Barcelo, Leopoldo Bertossi, Mikäel Monet. “On the Complexity of SHAP-Score-Based Explanations: Tractability via Knowledge Compilation and Non-Approximability Results”. *Journal of Machine Learning Research*, 2023, 24(63):1-58.
- 2 Leopoldo Bertossi and Jorge E. León. “Efficient Computation of Shap Explanation Scores for Neural Network Classifiers via Knowledge Compilation”. Proc. of JELIA’23, Springer LNCS 14281, 2023, pp. 49-64.

5.8 Circuits for Query Evaluation over Trees

Pierre Bourhis (CNRS – CRISTAL, Lille, FR)

License © Creative Commons BY 4.0 International license
© Pierre Bourhis

Querying trees via Tree automata presents a lot of interest because several important questions can be executed with a guaranteed efficient time. Over the last decades, different approaches have been presented to solve major query answering questions such as enumeration, probabilistic evaluation... In this survey, we review a particular approach which can be adapted to all these questions: a knowledge compilation approach. We present the different results that can be resolved by this approach and also its limits.

References

- 1 Antoine Amarilli, Pierre Bourhis, Florent Capelli, Mikäel Monet: Ranked Enumeration for MSO on Trees via Knowledge Compilation. CoRR abs/2310.00731 (2023)
- 2 Antoine Amarilli, Pierre Bourhis, Stefan Mengel, Matthias Niewerth: Enumeration on Trees with Tractable Combined Complexity and Efficient Updates. PODS 2019: 89-103
- 3 Antoine Amarilli, Pierre Bourhis, Stefan Mengel: Enumeration on Trees under Relabelings. ICDT 2018: 5:1-5:18
- 4 Antoine Amarilli, Pierre Bourhis, Louis Jachiet, Stefan Mengel: A Circuit-Based Approach to Efficient Enumeration. ICALP 2017: 111:1-111:15
- 5 Antoine Amarilli, Pierre Bourhis, Pierre Senellart: Provenance Circuits for Trees and Treelike Instances. ICALP (2) 2015: 56-68

5.9 From Queries to Circuits

Florent Capelli (University of Artois/CNRS – Lens, FR)

License © Creative Commons BY 4.0 International license
© Florent Capelli

Main reference Florent Capelli, Oliver Irwin: “Direct Access for Conjunctive Queries with Negations”, in Proc. of the 27th International Conference on Database Theory, ICDT 2024, March 25-28, 2024, Paestum, Italy, LIPIcs, Vol. 290, pp. 13:1–13:20, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024.

URL <https://doi.org/10.4230/LIPICS.ICDT.2024.13>

In this talk, we will review two algorithms to construct tractable circuits from a conjunctive query and a database whose size can be bounded using fractional hypertree width. The first one is a classical bottom up dynamic programming on a join tree of the conjunctive query, which can be seen as a generalization of Yannakakis approach. The second one is based on a top-down approach akin to exhaustive DPLL, an algorithm originally devised for solving #SAT. We will show that both algorithms construct very similar circuits on conjunctive queries but that DPLL can be applied to a more general setting without changing much of its structure.

5.10 Unified Reverse Data Management

Wolfgang Gatterbauer (Northeastern University – Boston, US)

License © Creative Commons BY 4.0 International license
© Wolfgang Gatterbauer

Joint work of Wolfgang Gatterbauer, Neha Makhija

What is a minimal set of tuples to delete from a database in order to eliminate all query answers? This problem is called “the resilience of a query” and is one of the key algorithmic problems underlying various forms of reverse data management, such as view maintenance, deletion propagation and causal responsibility. A long-open question is determining the conjunctive queries (CQs) for which resilience can be solved in PTIME.

We shed new light on this problem by proposing a unified Integer Linear Programming (ILP) formulation. It is unified in that it can solve both previously studied restrictions (e.g., self-join-free CQs under set semantics that allow a PTIME solution) and new cases (all CQs under set or bag semantics). It is also unified in that all queries and all database instances are treated with the same approach, yet the algorithm is guaranteed to terminate in PTIME for all known PTIME cases. In particular, we prove that for all known easy cases, the optimal solution to our ILP is identical to a simpler Linear Programming (LP) relaxation, which implies that standard ILP solvers return the optimal solution to the original ILP in PTIME.

In broader terms, we believe that using one single algorithm that can solve all queries (easy and hard) and then proving that it terminates in PTIME for the subset of PTIME queries will become a conventional and unified approach for attacking several other open problems in reverse data management.

References

- 1 Makhija and Gatterbauer. “A Unified Approach for Resilience and Causal Responsibility with Integer Linear Programming (ILP) and LP Relaxations”, SIGMOD 2024. <https://dl.acm.org/doi/pdf/10.1145/3626715>
- 2 Makhija and Gatterbauer. “Towards a Dichotomy for Minimally Factorizing the Provenance of Self-Join Free Conjunctive Queries”, PODS 2024. <https://arxiv.org/pdf/2105.14307>

5.11 Graph Explainability and Shapley

Floris Geerts (University of Antwerp, BE)

License © Creative Commons BY 4.0 International license
© Floris Geerts

Main reference Shichang Zhang, Yozen Liu, Neil Shah, Yizhou Sun: “GStarX: Explaining Graph Neural Networks with Structure-Aware Cooperative Games”, in Proc. of the Advances in Neural Information Processing Systems, Vol. 35, pp. 19810–19823, Curran Associates, Inc., 2022.

URL https://proceedings.neurips.cc/paper_files/paper/2022/file/7d53575463291ea6b5a23cf6e571f59b-Paper-Conference.pdf

Graph explainability is a critical aspect in understanding and interpreting complex relationships within graph-structured data. The need for transparent and interpretable models has led to the exploration of various methodologies, with a focus on providing insights into the contribution of individual nodes or edges in a graph. Shapley values, inspired by cooperative game theory, offer a principled approach to attribute values to each node, reflecting their marginal contributions to different coalitions. Myerson value further refines this concept by considering the externalities of a coalition, providing a more comprehensive understanding of node importance. In the context of graph explainability, Hamiache and Navarro score introduces a novel perspective by evaluating the relevance of nodes based on the information flow and connectivity patterns, offering a nuanced interpretation of their impact on the overall graph structure. Together, these approaches contribute to the development of explainable graph models, enabling stakeholders to gain deeper insights into the dynamics and significance of individual elements within complex graph data.

5.12 Lessons Learned from Building Systems for Provenance and Explanations

Boris Glavic (University of Illinois – Chicago, US)

License © Creative Commons BY 4.0 International license
© Boris Glavic

In this talk I will introduce to the audience lessons learned from my work and other group’s work on building systems for capturing and managing provenance and explanations. For instance, an underappreciated concept in developing such systems is that provenance creates a separate information flow in the system that does not conform to the standard way of how data flows through the operators of a query.

References

- 1 <https://vldb.org/pvldb/vol115/p451-niu.pdf>
- 2 <https://arxiv.org/pdf/1804.07156.pdf>
- 3 <http://sites.computer.org/debull/A18mar/p51.pdf>
- 4 <http://www.vldb.org/pvldb/vol113/p912-lee.pdf>
- 5 <https://dl.acm.org/doi/pdf/10.1145/3555041.3589731>
- 6 <https://inria.hal.science/hal-01851538/document>
- 7 <https://dl.acm.org/doi/pdf/10.14778/2824032.2824089>

5.13 Answering Database Queries Using Direct-Access Structures

Benny Kimelfeld (Technion – Haifa, IL)

License  Creative Commons BY 4.0 International license
© Benny Kimelfeld


The talk will describe recent results on the fine-grained complexity of database queries that involve joins, grouping, aggregation, and ordering. For some common aggregate functions (e.g., min, max, count, sum), such a query can be phrased as an ordinary conjunctive query over a database annotated with a suitable commutative semiring. I will discuss the ability to evaluate such queries by constructing, in quasilinear time in the database size (i.e., roughly the time it takes to read the database), a data structure that provides logarithmic-time direct access to the answers, ordered by a desired lexicographic order. This task is nontrivial since the number of answers might be larger than quasilinear in the database size, so, the data structure needs to provide a representation that is compact, easy to construct, and fast to access. The results provide classifications of queries, orderings, and semirings by the feasibility of such complexity guarantees.

References

- 1 Nofar Carmeli, Nikolaos Tziavelis, Wolfgang Gatterbauer, Benny Kimelfeld, Mirek Riedewald: Tractable Orders for Direct Access to Ranked Answers of Conjunctive Queries. PODS 2021: 325-341
- 2 Idan Eldar, Nofar Carmeli, Benny Kimelfeld: Direct Access for Answers to Conjunctive Queries with Aggregation. CoRR abs/2303.05327 (2023). ICDT 2024.

5.14 The Shapley Value in Database Management

Ester Livshits (University of Edinburgh, GB)

License  Creative Commons BY 4.0 International license
© Ester Livshits

Joint work of Ester Livshits, Leopoldo E. Bertossi, Benny Kimelfeld, Moshe Sebag
Main reference Ester Livshits, Leopoldo E. Bertossi, Benny Kimelfeld, Moshe Sebag: “The Shapley Value of Tuples in Query Answering”, *Log. Methods Comput. Sci.*, Vol. 17(3), 2021.
URL [https://doi.org/10.46298/LMCS-17\(3:22\)2021](https://doi.org/10.46298/LMCS-17(3:22)2021)

We consider two situations where we wish to quantify the responsibility of individual database tuples to the outcome. The first is query answering, where we wish to provide an explanation as to why we obtained a specific answer. The second is database inconsistency, where the goal is to identify the most problematic tuples. Some tuples may contribute more than others to the outcome, which can be a bit in the case of a Boolean query, a tuple or a number for conjunctive and aggregate queries, respectively, or a number indicating how inconsistent the database is (i.e., an inconsistency measure). To quantify the contribution of tuples, we use the well-known Shapley value that was introduced in cooperative game theory in the 1950s and has found applications in a plethora of domains. We investigate the applicability of the Shapley value in the two settings, as well as the computational aspects of its calculation in terms of complexity, algorithms, and approximation.

References

- 1 Ester Livshits, Leopoldo E. Bertossi, Benny Kimelfeld, and Moshe Sebag. “The Shapley Value of Tuples in Query Answering”. *Logical Methods in Computer Science* (2021). <https://lmcs.episciences.org/8437>

- 2 Ester Livshits and Benny Kimelfeld. “The Shapley Value of Inconsistency Measures for Functional Dependencies”. *Logical Methods in Computer Science* (2022). <https://lmcs.episciences.org/9705>

5.15 Provenance in Queries, Games, and Argumentation: Time for a Family Reunion

Bertram Ludäscher (University of Illinois at Urbana-Champaign, US)

License © Creative Commons BY 4.0 International license
© Bertram Ludäscher

Joint work of Bertram Ludäscher, Shawn Bowers, Yilin Xia

Main reference Bertram Ludäscher, Shawn Bowers, Yilin Xia: “Games, Queries, and Argumentation Frameworks: Towards a Family Reunion”, in Proc. of the 7th Workshop on Advances in Argumentation in Artificial Intelligence (AI³ 2023) co-located with the 22nd International Conference of the Italian Association for Artificial Intelligence (AIXIA 2023), Rome, Italy, November 9, 2023, CEUR Workshop Proceedings, Vol. 3546, CEUR-WS.org, 2023.

URL <https://ceur-ws.org/Vol-3546/paper06.pdf>

Consider the non-stratified, recursive query Q : $\text{win}(X) \text{ :- move}(X,Y)$, not $\text{win}(Y)$.

Its 2-valued reading states that a position x in a two-player game is won if there exists a move to a position y that is lost (not won). If the move graph contains cycles, drawn positions may occur (neither player can force a win). It is well known that the 3-valued well-founded semantics can be used to solve games: $\text{win}(x)$ is true, false, and undefined, respectively, iff x is won, lost, or drawn.

The query Q has been used in logic programming (e.g., to illustrate the well-founded semantics [3], in database theory (e.g., to show that stratified Datalog is strictly less expressive than the class of Fixpoint queries [2], and in formal argumentation (as a meta-interpreter for abstract argumentation frameworks).

Solved game graphs can be said to “explain themselves” (or contain their own provenance “for free”): The provenance of a won, lost, or drawn position is easily obtained via an RPQ-definable subgraph of the solved (labeled) game graph in which positions and moves have an associated value (won, lost, or drawn for positions, and winning, delaying, drawing, or blundering for moves, respectively). Since Q is a syntactic variant of Dung’s meta-interpreter for abstract argument frameworks AF [4], the provenance structure available in solved game graphs can be used to explain and justify the grounded (i.e., well-founded) extensions of AF. Another application of game provenance are query evaluation games: The n -ary version of Q can be understood as a normal form for Fixpoint, i.e., all Fixpoint queries can be rewritten into a game normal form, even when restricted to draw-free games [5]. For the subclass of FO queries (First-Order queries expressed in Datalog syntax), this normal form has been used to derive an elegant and powerful provenance representation that unifies how-provenance and why-not provenance [6].

References

- 1 B. Ludäscher, S. Bowers, and Y. Xia. Games, Queries, and Argumentation Frameworks: Towards a Family Reunion. *AI³@AI*IA* (2023)
- 2 P. Kolaitis, The expressive power of stratified logic programs, *Information and Computation* (1991).
- 3 A. Van Gelder, K. A. Ross, J. S. Schlipf, The Well-founded Semantics for General Logic Programs, *Journal of the ACM* (1991).
- 4 P. Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, *Logic Programming and n-Person Games, AI* (1995)

- 5 J. Flum, M. Kubierschky, and B. Ludäscher. Total and partial well-founded Datalog coincide. ICDT, Delphi. LNCS 1186 (1997)
- 6 S. Köhler, B. Ludäscher, and D. Zinn. First-Order Provenance Games. In Search of Elegance in the Theory and Practice of Computation (Peter Buneman Festschrift). LNCS 8000 (2013)

5.16 Impact of Self-Joins on Enumeration and Direct Access on Join Queries

Stefan Mengel (CNRS, CRIL – Lens, FR)

License  Creative Commons BY 4.0 International license
© Stefan Mengel

Joint work of Karl Bringmann, Nofar Carmeli, Stefan Mengel, Luc Segoufin

It has been known essentially since the introduction of conjunctive queries that self-joins have an impact on the evaluation of join queries. While in settings like answering Boolean queries and counting their complexity implications are completely understood, the situation is far less clear for other query answering tasks. In this talk, I will present some recent progress for enumeration (Carmeli and Segoufin 2023) and direct access (Bringmann, Carmeli, and Mengel 2023) showing that, even though these settings are often conceptually very close, self-joins behave very differently for them.

References

- 1 Karl Bringmann, Nofar Carmeli, Stefan Mengel: Tight Fine-Grained Bounds for Direct Access on Join Queries. CoRR abs/2201.02401 (2022)
- 2 Nofar Carmeli, Luc Segoufin: Conjunctive Queries With Self-Joins, Towards a Fine-Grained Enumeration Complexity Analysis. PODS 2023: 277-289

5.17 The Intensional-Extensional Problem in Probabilistic Databases

Mikaël Monet (INRIA Lille, FR)

License  Creative Commons BY 4.0 International license
© Mikaël Monet

Joint work of Mikaël Monet, Antoine Amarilli, Dan Suciu

Main reference Mikaël Monet: “Solving a Special Case of the Intensional vs Extensional Conjecture in Probabilistic Databases”, CoRR, Vol. abs/1912.11864, 2019.

URL <http://arxiv.org/abs/1912.11864>

Dalvi and Suciu established a dichotomy for probabilistic query evaluation (PQE) over tuple-independent databases, for unions of conjunctive queries (UCQs): for each UCQ, the problem is either solvable in PTIME, or is #P-hard. The UCQs for which the problem is in PTIME are called **safe**. Dalvi and Suciu’s algorithm on such a safe query relies essentially on the following three probabilistic rules: Independence, Negation, and Inclusion-Exclusion. In parallel, another method to obtain PTIME algorithms for PQE is through **knowledge compilation**: one first compiles the provenance of a query Q on a TID D as a Boolean circuit or diagram from the field of knowledge compilation (e.g., OBDDs, FBDDs, d-DNNFs, etc), and then uses this circuit to compute the probability. At a high-level, this type of algorithm makes use of the following three probabilistic rules: Independence, Negation, and **Disjoint union** (instead of inclusion-exclusion). This naturally leads to the following question, called the intensional-extensional problem: letting Q be a safe UCQ, can the tractability of PQE(Q) be captured with the knowledge compilation approach?

In this talk I will talk about this problem, present a technique that allowed to handle a specific class of UCQs, and discuss our ongoing work on the problem. In particular, I will present a neat combinatorial conjecture, that we named the “non-cancelling intersections” conjecture, that talks only about sets and the so-called Möbius function (i.e., no databases, no queries, no complexity). This talk is based on ongoing work with Antoine Amarilli, Louis Jachiet, and Dan Suciu.

References

- 1 Abhay Kumar Jha, Dan Suciu: Knowledge Compilation Meets Database Theory: Compiling Queries to Decision Diagrams. *Theory Comput. Syst.* 52(3): 403-440 (2013)
- 2 Michaël Monet: Solving a Special Case of the Intensional vs Extensional Conjecture in Probabilistic Databases. *PODS 2020*: 149-163

5.18 Revisiting Semiring Provenance for Datalog

Liat Peterfreund (The Hebrew University of Jerusalem, IL)

License © Creative Commons BY 4.0 International license
© Liat Peterfreund

Joint work of Camille Bourgaux, Pierre Bourhis, Liat Peterfreund, Michaël Thomazo

Main reference Camille Bourgaux, Pierre Bourhis, Liat Peterfreund, Michaël Thomazo: “Revisiting Semiring Provenance for Datalog”, in *Proc. of the 19th International Conference on Principles of Knowledge Representation and Reasoning*, pp. 91–101, 2022.

URL <https://doi.org/10.24963/kr.2022/10>

While the definition of semiring provenance is uncontroversial for unions of conjunctive queries, the picture is less clear for Datalog. Indeed, the original definition might include infinite computations and is not consistent with other proposals for Datalog semantics over annotated data. In this work, we propose and investigate several provenance semantics, based on different approaches for defining classical Datalog semantics. We study the relationship between these semantics, and introduce properties that allow us to analyze and compare them.

5.19 Datalog over (Pre-)Semirings

Reinhard Pichler (TU Wien, AT)

License © Creative Commons BY 4.0 International license
© Reinhard Pichler

Joint work of Mahmoud Abo Khamis, Hung Q. Ngo, Reinhard Pichler, Dan Suciu, Yisu Remy Wang

Main reference Mahmoud Abo Khamis, Hung Q. Ngo, Reinhard Pichler, Dan Suciu, Yisu Remy Wang: “Convergence of Datalog over (Pre-) Semirings”, in *Proc. of the PODS ’22: International Conference on Management of Data*, Philadelphia, PA, USA, June 12 – 17, 2022, pp. 105–117, ACM, 2022.

URL <https://doi.org/10.1145/3517804.3524140>

Datalog is a successful query language that extends relational calculus by recursion, has an elegant declarative semantics as well as a simple operational semantics, and admits several powerful optimizations such as semi-naïve evaluation and magic set rewriting. However, datalog also has its limitations since it only supports monotone queries over sets. This means, for instance, that aggregates (which are crucial in many data analytics tasks but are not monotone under set inclusion) are not supported in pure datalog.

In a seminal paper by Green, Karvounarakis, and Tannen [1], K-relations were introduced as a generalization of standard relations. In a K-relation, tuples are mapped to some semiring K. We can then consider standard relations as K-relations over the Boolean semiring, bags

of tuples as K-relations over the natural numbers, sparse tensors as K-relations over the reals, etc. Also provenance information at various levels of detail can be captured by an appropriate choice of the semiring K.

In this talk, I have presented our recent work [2, 3] on the query language datalog_o, which is based on the concept of K-relations and generalizes datalog to (pre-)semirings. In particular, I have shown how it can capture various computations involving aggregates as well as provenance information. Moreover, I have briefly mentioned convergence properties of datalog_o and some optimization techniques.

References

- 1 Todd J. Green, Gregory Karvounarakis, Val Tannen: Provenance semirings. PODS 2007: 31-40: <https://dl.acm.org/doi/10.1145/1265530.1265535>.
- 2 Mahmoud Abo Khamis, Hung Q. Ngo, Reinhard Pichler, Dan Suciu, Yisu Remy Wang: Convergence of Datalog over (Pre-) Semirings. PODS 2022: 105-117: <https://dl.acm.org/doi/10.1145/3517804.3524140>, full version (to appear in J.ACM): <https://arxiv.org/abs/2105.14435>.
- 3 Yisu Remy Wang, Mahmoud Abo Khamis, Hung Q. Ngo, Reinhard Pichler, Dan Suciu: Optimizing Recursive Queries with Program Synthesis. SIGMOD Conference 2022: 79-93: <https://dl.acm.org/doi/10.1145/3514221.3517827>.

5.20 MSO Enumeration over Words and their Representations

Cristian Riveros (PUC – Santiago de Chile, CL)

License  Creative Commons BY 4.0 International license
© Cristian Riveros

I will present a survey of MSO enumeration problems over words based on the model of annotated automata, a model for encoding MSO queries with output. In the first half, I will present the basic MSO enumeration problem and the representations needed for efficient enumeration. In the second half, I will go through extensions of this MSO enumeration problem, with the required extensions on the representations. Toward the end, I will present some open problems.

References

- 1 Martín Muñoz, Cristian Riveros: Constant-Delay Enumeration for SLP-Compressed Documents. ICDT 2023.
- 2 Martín Muñoz, Cristian Riveros: Streaming Enumeration on Nested Documents. ICDT 2022.
- 3 Antoine Amarilli, Pierre Bourhis, Louis Jachiet, Stefan Mengel: A Circuit-Based Approach to Efficient Enumeration. ICALP 2017.

5.21 Explanations for Aggregate Query Answers – An Overview

Sudeepa Roy (Duke University – Durham, US)

License © Creative Commons BY 4.0 International license
© Sudeepa Roy

Joint work of Michael Cafarella, Sainyam Galhotra, Boris Glavic, Amir Gilad, Chenjie Li, Zhengjie Miao, Sudeepa Roy, Babak Salimi, Dan Suciu, Brit Youngmann, Qitian Zeng

I will give an overview of different types of explanations for aggregate query answers answering user questions like why a value is high/low or higher/lower than another value. I will discuss explanations by intervention, counterbalance, augmented provenance, causal explanations, and actionable explanations. Explanations by Shapley Value will be covered in other talks.

References

- 1 Sudeepa Roy, Dan Suciu: A Formal Approach to Finding Explanations for Database Queries, SIGMOD 2014
- 2 Zhengjie Miao, Qitian Zeng, Boris Glavic, Sudeepa Roy: Going Beyond Provenance: Explaining Query Answers with Pattern-based Counterbalances, SIGMOD 2019
- 3 Chenjie Li, Zhengjie Miao, Qitian Zeng, Boris Glavic, Sudeepa Roy: Putting Things into Context: Rich Explanations for Query Answers using Join Graphs, SIGMOD 2021
- 4 Sainyam Galhotra, Amir Gilad, Sudeepa Roy, Babak Salimi: Hyper: Hypothetical Reasoning With What-If and How-To Queries Using a Probabilistic Causal Approach, SIGMOD 2022
- 5 Brit Youngmann, Amir Gilad, and Michael Cafarella, Sudeepa Roy: Summarized Causal Explanations For Aggregate Views, SIGMOD 2024

5.22 Training Invariant Machine Learning Models with Incomplete Data

Babak Salimi (University of California, San Diego – La Jolla, US)

License © Creative Commons BY 4.0 International license
© Babak Salimi

Main reference Jiongli Zhu, Sainyam Galhotra, Nazanin Sabri, Babak Salimi: “Consistent Range Approximation for Fair Predictive Modeling”, Proc. VLDB Endow., Vol. 16(11), pp. 2925–2938, 2023.

URL <https://doi.org/10.14778/3611479.3611498>

In this talk, I aim to discuss the significant challenge of learning machine learning models that satisfy invariant properties under conditional independence constraints. The importance of this problem will be illustrated through various real-world examples, emphasizing its relevance and urgency. Subsequently, I will analyze existing approaches and their shortcomings, especially in situations where data is compromised by quality issues such as selection bias. To overcome these obstacles, I will introduce a framework inspired by techniques for querying incomplete data in data management. This framework is tailored to effectively handle the specific challenges posed by incomplete datasets. Additionally, I will demonstrate its application in the context of algorithmic fairness.

5.23 Expected Shapley-Like Scores of Boolean Functions: Complexity and Applications to Probabilistic Databases

Pierre Senellart (ENS, PSL University – Paris, FR)

License  Creative Commons BY 4.0 International license
© Pierre Senellart

Joint work of Pratik Karmakar, Mikaël Monet, Pierre Senellart, Stephane Bressan

Main reference Pratik Karmakar, Mikaël Monet, Pierre Senellart, Stephane Bressan: “Expected Shapley-Like Scores of Boolean functions: Complexity and Applications to Probabilistic Databases”, Proc. ACM Manag. Data, Vol. 2(2), Association for Computing Machinery, 2024.

URL <https://doi.org/10.1145/3651593>

Shapley values, originating in game theory and increasingly prominent in explainable AI, have been proposed to assess the contribution of facts in query answering over databases, along with other similar power indices such as Banzhaf values. In this work we adapt these Shapley-like scores to probabilistic settings, the objective being to compute their expected value. We show that the computations of expected Shapley values and of the expected values of Boolean functions are interreducible in polynomial time, thus obtaining the same tractability landscape. We investigate the specific tractable case where Boolean functions are represented as deterministic decomposable circuits, designing a polynomial-time algorithm for this setting. We present applications to probabilistic databases through database provenance, and an effective implementation of this algorithm within the ProVSQL system, which experimentally validates its feasibility over a standard benchmark.

References

- 1 Pratik Karmakar, Mikaël Monet, Pierre Senellart, and Stéphane Bressan, 2024. Expected Shapley-Like Scores of Boolean Functions: Complexity and Applications to Probabilistic Databases, <https://arxiv.org/abs/2401.06493>
- 2 Daniel Deutch, Nave Frost, Benny Kimelfeld, and Mikaël Monet. 2022. Computing the Shapley value of facts in query answering. In SIGMOD Conference. ACM, 1570–1583.
- 3 Pierre Senellart, Louis Jachiet, Silviu Maniu, and Yann Ramusat. 2018. ProVSQL: Provenance and Probability Management in PostgreSQL. Proc. VLDB Endow. 11, 12 (2018), 2034–2037.

5.24 The Importance of Parameters in Database Queries

Christoph Standke (RWTH Aachen, DE)

License  Creative Commons BY 4.0 International license
© Christoph Standke

Joint work of Martin Grohe, Benny Kimelfeld, Peter Lindner, Christoph Standke

Main reference Martin Grohe, Benny Kimelfeld, Peter Lindner, Christoph Standke: “The Importance of Parameters in Database Queries”, in Proc. of the 27th International Conference on Database Theory, ICDT 2024, March 25-28, 2024, Paestum, Italy, LIPIcs, Vol. 290, pp. 14:1–14:17, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024.

URL <https://doi.org/10.4230/LIPICS.ICDT.2024.14>

In this talk, I will introduce a framework for quantifying the importance of the choices of parameter values to the result of a query over a database. In our framework, the importance of a parameter is its SHAP score and we make the case for the rationale of using this score by showing that we arrive at this score in two different, apparently opposing, approaches to quantifying the contribution of a parameter. We then point out that this framework yields an interesting complexity-theoretic landscape.

5.25 From Shapley Value to Model Counting and Back

Dan Suciu (*University of Washington – Seattle, US*)

License  Creative Commons BY 4.0 International license
© Dan Suciu

We study the problem of quantifying the contribution of each Boolean variable to the satisfying assignments of a Boolean function, based on the Shapley value. This problem was introduced by Livshits et al. in order to quantify the contribution of an input tuple to the output of a query. We prove polynomial-time equivalence between computing Shapley values and model counting, for classes of Boolean functions that are closed under substitutions of variables with disjunctions of fresh variables. This result settles an open problem raised by Deutch et al., which sought to connect the Shapley value computation to probabilistic query evaluation.

References

- 1 Ester Livshits, Leopoldo E. Bertossi, Benny Kimelfeld, Moshe Sebag: The Shapley Value of Tuples in Query Answering. *Log. Methods Comput. Sci.* 17(3) (2021)
- 2 Daniel Deutch, Nave Frost, Benny Kimelfeld, Mikaël Monet: Computing the Shapley Value of Facts in Query Answering. *SIGMOD Conference 2022*: 1570-1583
- 3 Ahmet Kara, Dan Olteanu, Dan Suciu: From Shapley Value to Model Counting and Back. *CoRR abs/2306.14211* (2023) (To appear in *PODS'2024*)

5.26 Answering Quantile Join Queries by Representing Inequality Predicates Efficiently

Nikolaos Tziavelis (*Northeastern University – Boston, US*)

License  Creative Commons BY 4.0 International license
© Nikolaos Tziavelis

Joint work of Nikolaos Tziavelis, Nofar Carmeli, Wolfgang Gatterbauer, Benny Kimelfeld, Mirek Riedewald
Main reference Nikolaos Tziavelis, Nofar Carmeli, Wolfgang Gatterbauer, Benny Kimelfeld, Mirek Riedewald: “Efficient Computation of Quantiles over Joins”, in *Proc. of the 42nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2023, Seattle, WA, USA, June 18-23, 2023*, pp. 303–315, ACM, 2023.
URL <https://doi.org/10.1145/3584372.3588670>


We consider the complexity of answering Quantile Join Queries, which ask for the answer at a specified relative position (e.g., 50% for the median) under some ordering over the answers to an ordinary Join Query (JQ). Compared to the task of direct access, this task is easier since only one access is required. The goal is to avoid materializing the set of all join answers, and to achieve quasilinear time in the size of the database, regardless of the total number of answers. The tractability of such a query does not only depend on the join structure, but also on the desired order. We show an algorithm that covers all known tractable cases by iteratively using a “trimming” subroutine which removes query answers that are higher or lower (according to the ranking function) than a certain answer determined as the “pivot”. Trimming essentially adds inequality predicates to our initial query and an efficient representation of these inequalities implies efficient Quantile Join Query answering for a large class of ranking functions.

6 Open problems

6.1 A problem on unambiguous DNFs

Mikaël Monet (INRIA Lille, FR)

License  Creative Commons BY 4.0 International license

 Mikaël Monet


URL <https://cstheory.stackexchange.com/q/53733>

I presented the problem that can be found here: <https://cstheory.stackexchange.com/q/53733>.

6.2 A question about representability of probabilistic databases

Christoph Standke (RWTH Aachen, DE)

License  Creative Commons BY 4.0 International license

 Christoph Standke

Joint work of Christoph Standke, Peter Lindner, Dan Suciu, Dan Olteanu, Christopher Ré, Christoph Koch
Main reference Dan Suciu, Dan Olteanu, Christopher Ré, Christoph Koch: “Probabilistic Databases”, Morgan & Claypool Publishers, 2011.

URL <https://doi.org/10.2200/S00362ED1V01Y201105DTM016>

Given a finite probabilistic database as a set of instance-probability pairs, (how) can we decide whether this probabilistic database can be obtained via a finite tuple-independent probabilistic database and a view consisting of conjunctive queries?

Participants

- Antoine Amarilli
Telecom Paris, FR
- Albert Atserias
UPC Barcelona Tech, ES
- Pablo Barcelo
PUC – Santiago de Chile, CL
- Christoph Berkholz
TU Ilmenau, DE
- Leopoldo Bertossi
SKEMA Business School –
Montréal, CA
- Pierre Bourhis
CNRS – CRISStAL, Lille, FR
- Florent Capelli
University of Artois/CNRS –
Lens, FR
- Wolfgang Gatterbauer
Northeastern University –
Boston, US
- Floris Geerts
University of Antwerp, BE
- Amir Gilad
The Hebrew University of
Jerusalem, IL
- Boris Glavic
University of Illinois –
Chicago, US
- Benny Kimelfeld
Technion – Haifa, IL
- Ester Livshits
University of Edinburgh, GB
- Bertram Ludäscher
University of Illinois at
Urbana-Champaign, US
- Stefan Mengel
CNRS, CRIL – Lens, FR
- Mikaël Monet
INRIA Lille, FR
- Liat Peterfreund
The Hebrew University of
Jerusalem, IL
- Reinhard Pichler
TU Wien, AT
- Cristian Riveros
PUC – Santiago de Chile, CL
- Sudeepa Roy
Duke University – Durham, US
- Babak Salimi
University of California,
San Diego – La Jolla, US
- Pierre Senellart
ENS, PSL University – Paris, FR
- Christoph Standke
RWTH Aachen, DE
- Dan Suciu
University of Washington –
Seattle, US
- Nikolaos Tziavelis
Northeastern University –
Boston, US
- Harry Vinall-Smeeth
TU Ilmenau, DE



Symmetric Cryptography

Christof Beierle^{*1}, Bart Mennink^{*2}, María Naya-Plasencia^{*3},
Yu Sasaki^{*4}, and Rachelle Heim Boissier^{†5}

- 1 Ruhr-Universität Bochum, DE. christof.beierle@rub.de
- 2 Radboud University Nijmegen, NL. b.mennink@cs.ru.nl
- 3 INRIA – Paris, FR. maria.naya_plasencia@inria.fr
- 4 NTT – Tokyo, JP. yu.sasaki.sk@hco.ntt.co.jp
- 5 UVSQ, Paris Saclay, FR. rachelle.heim@uvsq.fr

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar “Symmetric Cryptography” (24041). The seminar was held on January 21–26, 2024 in Schloss Dagstuhl – Leibniz Center for Informatics. This was the ninth seminar in the series “Symmetric Cryptography”. Previous editions were held in 2007, 2009, 2012, 2014, 2016, 2018, 2020 and 2022. Participants of the seminar presented their ongoing work and new results on topics of cryptanalysis and (post-quantum) provable security of symmetric cryptographic primitives. Participants also worked together within seven research group dedicated to various topics (Cryptanalysis of Poseidon, Cryptanalysis of TEA-3, Exploitation of the wrong key randomization hypothesis non-conformity in key recovery attacks, Cryptanalysis of SCARF, Differential cryptanalysis and more, Key control security and Security of sponge combiners). In this report, a brief summary of the seminar is given, followed by the abstracts of given talks and a summary of the progress of each research group.

Seminar January 21–26, 2024 – <https://www.dagstuhl.de/24041>

2012 ACM Subject Classification Security and privacy → Cryptanalysis and other attacks;
Security and privacy → Symmetric cryptography and hash functions

Keywords and phrases Lightweight Cryptography, New Applications of Symmetric Cryptography, Permutation-Based Cryptography

Digital Object Identifier 10.4230/DagRep.14.1.72

1 Executive Summary

Christof Beierle (Ruhr-Universität Bochum, DE, christof.beierle@rub.de)

Bart Mennink (Radboud University Nijmegen, NL, b.mennink@cs.ru.nl)

María Naya-Plasencia (INRIA – Paris, FR, maria.naya_plasencia@inria.fr)

Yu Sasaki (NTT – Tokyo, JP, yu.sasaki.sk@hco.ntt.co.jp)

License © Creative Commons BY 4.0 International license
© Christof Beierle, Bart Mennink, María Naya-Plasencia, and Yu Sasaki

IT Security plays an increasingly crucial role in our everyday life and business. Virtually all modern security solutions are based on cryptographic primitives. Symmetric cryptography deals with the case where both the sender and the receiver of a message are using the same key. Due to their good performance, symmetric cryptosystems are the main workhorses of cryptography and are highly relevant not only for academia, but also for industrial activities.

* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Symmetric Cryptography, *Dagstuhl Reports*, Vol. 14, Issue 1, pp. 72–89

Editors: Christof Beierle, Bart Mennink, María Naya-Plasencia, and Yu Sasaki



DAGSTUHL
REPORTS

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

For this Dagstuhl Seminar we focused on several topics, which we believe to be of great importance for the research community and, likewise, to have a positive impact on industry and the deployment of secure crypto in the future.

- **Follow Up on Main Results from Last Dagstuhl Seminar.** At the last Dagstuhl Seminar on symmetric cryptography in 2022, the participants were divided into six groups in order to discuss research topics proposed by each participant. The discussions were very productive and there were and will be publications from several groups. We believe that the discussions and results from these 2022 work groups reflect the main interests of the community and are useful topics to continue to discuss at the Dagstuhl Seminar in 2024. Participants at the 2024 Dagstuhl Seminar who also participated in the work groups in 2022 were thus invited to present their finished results.
- **Design and Analysis of Symmetric Crypto for New Applications.** Recently, the design of symmetric-key primitives has started to focus on different types of optimization. Those optimizations could be with respect to performance and with respect to special security requirements. Stated differently, one first considers a target application (such as multi-party computation or non-interactive zero-knowledge proofs), and only then designs symmetric-key primitives for this purpose. This causes a paradigm shift in design criteria. During this seminar, we explored the security of recently introduced ciphers that were designed specifically for such target applications, and develop novel ciphers with improved security arguments and guarantees.
- **Generic Analysis of Emerging Modes.** Permutation-based cryptography has gained astounding popularity in the last decade, and security proofs are performed in the ideal permutation model. A similar phenomenon is visible in various ideal cipher-based constructions that have appeared recently. In this seminar, we explored how results with different models (such as a standard model and an ideal model) compare from a theoretical perspective, and investigated what cryptanalytical results on certain primitives mean for the targeted construction.

Seminar Program

The seminar program consisted of short presentations and group meetings. Presentations were about the above topics and other relevant areas of symmetric cryptography, including state-of-the-art cryptanalytic techniques and new designs. The list of abstracts for talks given during the seminar can be found below. Also, participants met in smaller groups and spent a significant portion of the week, each group intensively discussing a specific research topic. There were seven research groups:

- Cryptanalysis of Poseidon;
- Cryptanalysis of TEA-3;
- Exploitation of the wrong key randomization hypothesis non-conformity in key recovery attacks;
- Cryptanalysis of SCARF;
- Differential cryptanalysis and more;
- Key control security;
- Security of sponge combiners.

On the last day of the week the leaders of each group gave brief summaries of achievements. An abstract corresponding to each research group can be found below. Some teams continued working on the topic after the seminar and started new research collaborations.

2 Table of Contents

Executive Summary

Christof Beierle, Bart Mennink, María Naya-Plasencia, and Yu Sasaki 72

Overview of Talks

Follow-up on Differential Meet-In-The-Middle Cryptanalysis <i>Zahra Ahmadian</i>	76
A New Post-Quantum Proof Framework <i>Ritam Bhaumik</i>	76
A generic algorithm for efficient key recovery in differential attacks – and its associated tool <i>Christina Boura, Nicolas David, Patrick Derbez, Rachelle Heim Boissier, and María Naya-Plasencia</i>	77
Differential Meet-In-The-Middle Cryptanalysis <i>Christina Boura</i>	77
Generalized Initialization of the Duplex Construction <i>Christoph Dobraunig and Bart Mennink</i>	78
Key Control Security of PRF-Based KDFs; Introduction and Preliminary Cryptanalysis Results <i>Tetsu Iwata</i>	78
Range-Restricted Vertex Labeling and Its Applications <i>Ashwin Jha</i>	78
On Boomerang Attacks on Quadratic Feistel Ciphers <i>Virginie Lallemand</i>	79
Algebraic Attack on FHE-Friendly Cipher HERA Using Multiple Collisions <i>Willi Meier</i>	79
Revisiting the Indifferentiability of the Sum of Permutations <i>Bart Mennink</i>	80
Revisiting Vector-input MACs <i>Kazuhiko Minematsu</i>	80
On INT-RUP security analysis of GCM <i>Akiko Inoue, Tetsu Iwata, and Kazuhiko Minematsu</i>	81
Indifferentiability of 6-round Feistel <i>Mridul Nandi</i>	81
The Algebraic Freelunch: Efficient Gröbner Basis Attacks Against Arithmetization-Oriented Primitives <i>Léo Perrin</i>	82
The t-wise Independence of SPNs <i>Stefano Tessaro</i>	83

Working groups

First results on the multivariate cryptanalysis of Poseidon <i>Lorenzo Grassi, Antoine Joux, Patrick Neumann, Léo Perrin, Christian Rechberger, Ferdinand Sibleyras, Aleksei Udovenko, and Qingju Wang</i>	84
Cryptanalysis of TEA3 <i>Subhadeep Banik, Christof Beierle, Anne Canteaut, Patrick Felke, Nils Gregor Leander, Gaëtan Leurent, Yann Rotella, Sondre Rønjom, and Siwei Sun</i>	85
Exploitation of the Wrong Key Randomization Hypothesis Non-conformity in Key Recovery Attacks <i>Zhenzhen Bao and Nils Gregor Leander</i>	86
Cryptanalysis of SCARF <i>Christina Boura, Zahra Ahmadian, Yanis Belkheyar, Christoph Dobraunig, Henri Gilbert, Shahram Rasoolzadeh, Dhiman Saha, Tyge Tiessen, and Yosuke Todo</i>	86
Differential cryptanalysis and more <i>Patrick Derbez, Orr Dunkelman, Maria Eichlseder, Ryoma Ito, Virginie Lallemand, and María Naya-Plasencia</i>	87
Key Control Security Group <i>Tetsu Iwata, Ritam Bhaumik, Avijit Dutta, Akiko Inoue, Ashwin Jha, Kazuhiko Minematsu, Mridul Nandi, Yu Sasaki, Meltem Sonmez Turan, Stefano Tessaro, and Aishwarya Thiruvengadam</i>	87
Security of sponge combiners <i>Charlotte Lefevre, Rachele Heim Boissier, Bart Mennink, and Bart Preneel</i>	88
Participants	89

3 Overview of Talks

3.1 Follow-up on Differential Meet-In-The-Middle Cryptanalysis

Zahra Ahmadian (*Shahid Beheshti University – Tehran, IR*)

License © Creative Commons BY 4.0 International license
© Zahra Ahmadian

Joint work of Zahra Ahmadian, Akram Khalesi, Dounia M’Foukh, Hossein Moghimi, María Naya-Plasencia

In this presentation, we generalize the differential meet-in-the-middle attack proposed at Crypto 2023 [1] to incorporate truncated differentials. Subsequently, we propose three enhancements to the differential-MITM attack: a stronger parallel partitioning technique covering more rounds, probabilistic key recovery requiring less key material, and benefiting from the state-test technique previously proposed in the context of impossible differential attacks.

Using a MILP-based tool to automate the search for optimized overall complexity, incorporating some of the proposed improvements, we develop the best-known attacks on the cipher CRAFT, reaching 23 rounds compared to the previous 21. We also improve the best attack on the 25-round SKINNY-128-384 and provide a new attack on the 23-round SKINNY-64-192.

References

- 1 Boura, C., David, N., Derbez, P., Leander, G., Naya-Plasencia, M.: Differential meet-in-the-middle cryptanalysis. In: *Advances in Cryptology – CRYPTO 2023 – 43rd Annual International Cryptology Conference, CRYPTO 2023, Santa Barbara, CA, USA, August 20-24, 2023, Proceedings, Part III. Lecture Notes in Computer Science*, vol. 14083, pp. 240–272. Springer (2023)

3.2 A New Post-Quantum Proof Framework

Ritam Bhaumik (*EPFL – Lausanne, CH*)

License © Creative Commons BY 4.0 International license
© Ritam Bhaumik

Joint work of Ritam Bhaumik, Benoît Cogliati, Jordan Ethan, Ashwin Jha

Main reference Ritam Bhaumik, Benoît Cogliati, Jordan Ethan, Ashwin Jha: “On Quantum Secure Compressing Pseudorandom Functions”, in *Proc. of the Advances in Cryptology – ASIACRYPT 2023 – 29th International Conference on the Theory and Application of Cryptology and Information Security, Guangzhou, China, December 4-8, 2023, Proceedings, Part III, Lecture Notes in Computer Science*, Vol. 14440, pp. 34–66, Springer, 2023.

URL https://doi.org/10.1007/978-981-99-8727-6_2

While quantum attacks on symmetric cryptosystems have been less devastating than those on certain popular public-key cryptosystems, classical provable security results on symmetric modes are not trivial to extend to the Q2 setting where the adversary has superposition access to the mode. In this work we attempt to build a generic proof framework applicable to such games by building on several previous works on compressed oracles.

3.3 A generic algorithm for efficient key recovery in differential attacks – and its associated tool

Christina Boura (University of Versailles, FR), Nicolas David, Patrick Derbez (University of Rennes, FR), Rachele Heim Boissier (University of Versailles, FR), and María Naya-Plasencia (INRIA – Paris, FR)

License © Creative Commons BY 4.0 International license

© Christina Boura, Nicolas David, Patrick Derbez, Rachele Heim Boissier, and María Naya-Plasencia

Main reference Christina Boura, Nicolas David, Patrick Derbez, Rachele Heim Boissier, María Naya-Plasencia: “A generic algorithm for efficient key recovery in differential attacks – and its associated tool”, 2024.

URL <https://eprint.iacr.org/2024/288>

Differential cryptanalysis is an old and powerful attack against block ciphers. While different techniques have been introduced throughout the years to improve the complexity of this attack, the key recovery phase remains a tedious and error-prone procedure. In this talk, we present a new algorithm and its associated tool that permits, given a distinguisher, to output an efficient key guessing strategy. Our tool can be applied to SPN ciphers whose linear layer consists of a bit-permutation and whose key schedule is linear or almost linear. It can be used not only to help cryptanalysts find the best differential attack on a given cipher but also to assist designers in their security analysis. We applied our tool to four targets: RECTANGLE, PRESENT-80, SPEEDY-7-192 and GIFT-64. We extend the previous best attack on RECTANGLE-128 by one round and the previous best differential attack against PRESENT-80 by 2 rounds. We improve a previous key recovery step in an attack against SPEEDY and present more efficient key recovery strategies for RECTANGLE-80 and GIFT. Our tool outputs the results in only a second for most targets.

3.4 Differential Meet-In-The-Middle Cryptanalysis

Christina Boura (University of Versailles, FR)

License © Creative Commons BY 4.0 International license

© Christina Boura

Joint work of Christina Boura, Nicolas David, Patrick Derbez, Gregor Leander, María Naya-Plasencia

Main reference Christina Boura, Nicolas David, Patrick Derbez, Gregor Leander, María Naya-Plasencia: “Differential Meet-In-The-Middle Cryptanalysis”, in Proc. of the Advances in Cryptology – CRYPTO 2023 – 43rd Annual International Cryptology Conference, CRYPTO 2023, Santa Barbara, CA, USA, August 20-24, 2023, Proceedings, Part III, Lecture Notes in Computer Science, Vol. 14083, pp. 240–272, Springer, 2023.

URL https://doi.org/10.1007/978-3-031-38548-3_9

In this talk we introduce the differential meet-in-the-middle framework, a new cryptanalysis technique for symmetric primitives. Our new cryptanalysis method combines techniques from both meet-in-the-middle and differential cryptanalysis. As such, the introduced technique can be seen as a way of extending meet-in-the-middle attacks and their variants but also as a new way to perform the key recovery part in differential attacks. We apply our approach to SKINNY-128-384 in the single-key model and to AES-256 in the related-key model. Our attack on SKINNY-128-384 permits to break 25 out of the 56 rounds of this variant and improves by two rounds the previous best known attacks. For AES-256 we attack 12 rounds by considering two related keys, thus outperforming the previous best related-key attack on AES-256 with only two related keys by 2 rounds.

3.5 Generalized Initialization of the Duplex Construction

Christoph Dobraunig (Intel – Villach, AT) and Bart Mennink (Radboud University Nijmegen, NL)

License © Creative Commons BY 4.0 International license

© Christoph Dobraunig and Bart Mennink

Main reference Christoph Dobraunig, Bart Mennink: “Generalized Initialization of the Duplex Construction”, in Proc. of the Applied Cryptography and Network Security – 22nd International Conference, ACNS 2024, Abu Dhabi, United Arab Emirates, March 5-8, 2024, Proceedings, Part II, Lecture Notes in Computer Science, Vol. 14584, pp. 460–484, Springer, 2024.

URL https://doi.org/10.1007/978-3-031-54773-7_18

If we consider (authenticated) encryption schemes based on the sponge/duplex construction, it is typically assumed that the adversary has the capability to choose the nonce/IV on its will (except for repetitions). In this talk, we discuss how the security changes if restrictions on the choice of the nonce are imposed, varying from the global nonce case over the random nonce case to the nonce on key case.

3.6 Key Control Security of PRF-Based KDFs; Introduction and Preliminary Cryptanalysis Results

Tetsu Iwata (Nagoya University, JP)

License © Creative Commons BY 4.0 International license

© Tetsu Iwata

Joint work of Tetsu Iwata, Keisuke Ozeki

NIST SP 800-108r1 specifies Key Derivation Functions (KDFs) based on PseudoRandom Functions (PRFs). In the document, the key control security is discussed, and it is pointed out KDFs based on CMAC have security issues. In this talk, we review the notion of the key control security and the security issues. We then point out similar security issues are in other block cipher based PRFs. We also present an attempt to formalize a cryptographic definition of the key control security, and discuss possible directions for future research.

3.7 Range-Restricted Vertex Labeling and Its Applications

Ashwin Jha (Ruhr-Universität Bochum, DE)

License © Creative Commons BY 4.0 International license

© Ashwin Jha

Joint work of Ashwin Jha, Mridul Nandi, Abishanka Saha

Most of the beyond-the-birthday-bound deterministic MAC (or PRF) constructions, like PMAC+ and LightMAC+, can be viewed as an instance of the Double-block Hash-then-Sum (DbHtS) paradigm (DDNP, IACR ToSC 2018). It is well-known (KLL, EUROCRYPT 2020; JN, JoC 2020; LNS, CRYPTO 2018) that DbHtS constructions are secure up to roughly $2^{3n/4}$ queries, where n denotes the block size. In contrast, the security guarantees for single-keyed variants of DbHtS, namely PMAC+ and LightMAC+, have only been proven up to $2^{2n/3}$ queries, with no known matching attack. In this work, we revisit the security of single-keyed DbHtS and map the problem to a graph vertex labeling problem where the labels are to be chosen outside some prohibited set (a strict subset of $\{0, 1\}^n$). We derive a strong lower bound for the number of such valid vertex labelings, under certain randomness assumptions

on the prohibited set. This directly implies security up to $2^{3n/4}$ queries for single-keyed DbHtS. Furthermore, we show that the hash functions in the single-keyed variants of PMAC+ and LightMAC+ satisfy the required conditions. Consequently, the single-keyed PMAC+ and LightMAC+ are shown to be equivalent (up to some constant factors) to their multi-keyed counterparts in terms of security. We conclude with a discussion on the potential applications of our techniques to similar problems in the random permutation model.

3.8 On Boomerang Attacks on Quadratic Feistel Ciphers

Virginie Lallemand (LORIA – Nancy, FR)

License © Creative Commons BY 4.0 International license
© Virginie Lallemand

Joint work of Xavier Bonnetain, Virginie Lallemand

Main reference Xavier Bonnetain, Virginie Lallemand: “On Boomerang Attacks on Quadratic Feistel Ciphers New results on KATAN and Simon”, IACR Trans. Symmetric Cryptol., Vol. 2023(3), pp. 101–145, 2023.

URL <https://doi.org/10.46586/TOSC.V2023.I3.101-145>

We study the application of the boomerang attack technique to ciphers following a Feistel construction and having a quadratic round function. We prove that many previously published papers give highly inaccurate probability approximations of the distinguishers they use. We next propose a new SMT model that takes into account our findings and we are able to propose a 19-round distinguisher of Simon-32/64 that we convert into the (to the best of our knowledge) first 25-round attack.

3.9 Algebraic Attack on FHE-Friendly Cipher HERA Using Multiple Collisions

Willi Meier (FH Nordwestschweiz – Windisch, CH)

License © Creative Commons BY 4.0 International license
© Willi Meier

Joint work of Willi Meier, Fukang Liu, Abul Kalam, Santanu Sarkar

Main reference Fukang Liu, Abul Kalam, Santanu Sarkar, Willi Meier: “Algebraic Attack on FHE-Friendly Cipher HERA Using Multiple Collisions”, IACR Cryptol. ePrint Arch., p. 1800, 2023.

URL <https://eprint.iacr.org/2023/1800>

In this work, the first third-party cryptanalysis of the FHE-friendly stream cipher HERA is performed, by showing how to mount new algebraic attacks with multiple collisions in the round keys. Specifically, according to the special way to randomize the round keys in HERA, we peel off the last nonlinear layer by using collisions in the last-round key and a simple property of the power map. In this way, we construct an overdefined system of equations of a much lower degree in the key, and efficiently solve the system via the linearization technique. As a result, for HERA with 192 and 256 bits of security, respectively, we break some parameters under the same assumption made by designers that the algebra constant ω for Gaussian elimination is $\omega = 2$, *i.e.*, Gaussian elimination on an $n \times n$ matrix takes $\mathcal{O}(n^\omega)$ field operations. If using more conservative choices like $\omega \in 2.8, 3$, our attacks can also successfully reduce the security margins of some variants of HERA to only 1 round.

3.10 Revisiting the Indifferentiability of the Sum of Permutations

Bart Mennink (Radboud University Nijmegen, NL)


License  Creative Commons BY 4.0 International license
© Bart Mennink

Joint work of Aldo Gunsing, Ritam Bhaumik, Ashwin Jha, Bart Mennink, Yaobin Shen
Main reference Aldo Gunsing, Ritam Bhaumik, Ashwin Jha, Bart Mennink, Yaobin Shen: “Revisiting the Indifferentiability of the Sum of Permutations”, in Proc. of the Advances in Cryptology – CRYPTO 2023 – 43rd Annual International Cryptology Conference, CRYPTO 2023, Santa Barbara, CA, USA, August 20-24, 2023, Proceedings, Part III, Lecture Notes in Computer Science, Vol. 14083, pp. 628–660, Springer, 2023.
URL https://doi.org/10.1007/978-3-031-38548-3_21

The sum of two n -bit pseudorandom permutations is known to behave like a pseudorandom function with n bits of security. A recent line of research has investigated the security of two public n -bit permutations and its degree of indifferentiability. Mandal et al. (INDOCRYPT 2010) proved $2n/3$ -bit security, Mennink and Preneel (ACNS 2015) pointed out a non-trivial flaw in their analysis and re-proved $(2n/3 - \log_2(n))$ -bit security. Bhattacharya and Nandi (EUROCRYPT 2018) eventually improved the result to n -bit security. Recently, Gunsing at CRYPTO 2022 already observed that a proof technique used in this line of research only holds for sequential indifferentiability. We revisit the line of research in detail, and observe that the strongest bound of n -bit security has two other serious issues in the reasoning, the first one is actually the same non-trivial flaw that was present in the work of Mandal et al., while the second one discards biases in the randomness influenced by the distinguisher. More concretely, we introduce two attacks that show limited potential of different approaches. We (i) show that the latter issue that discards biases only holds up to $2^{3n/4}$ queries, and (ii) perform a differentiability attack against their simulator in $2^{5n/6}$ queries. On the upside, we revive the result of Mennink and Preneel and show $(2n/3 - \log_2(n))$ -bit regular indifferentiability security of the sum of public permutations.

3.11 Revisiting Vector-input MACs

Kazuhiko Minematsu (NEC – Kawasaki, JP)

License  Creative Commons BY 4.0 International license
© Kazuhiko Minematsu

Joint work of Kazuhiko Minematsu, Isamu Furuya

Rogaway and Shrimpton (RS06) presented the idea of vector-input MAC that accepts a vector consisting of variable-length bit strings. A vector-input MAC could be built on a conventional (bit) string-input MAC, e.g. CMAC, with an injective encoding. RS06 pointed out an efficiency loss in this method and presented a general construction S2V that is more efficient than the encoding-based method. However, despite its potential, their work on vector-input MAC has been largely overlooked for more than 16 years. We revisit RS06’s treatment of vector-input MAC and showed that the topic is more subtle than initially considered. We first formally define the problem of vector-input MAC and propose a natural efficiency goal for vector-input MACs as a counterpart of what has been considered for string-input MACs. Since S2V with any string-input MAC mode never achieves this efficiency goal, we propose a new MAC mode, VecMAC, that achieves this goal for any vector space. VecMAC is a variant of the popular PMAC. However, its use of tweaks is more involved and conceptually different from PMAC. We also provide preliminary implementation results.

3.12 On INT-RUP security analysis of GCM

Akiko Inoue (NEC – Kawasaki, JP), Tetsu Iwata (Nagoya University, JP), and Kazuhiko Minematsu (NEC – Kawasaki, JP)

License © Creative Commons BY 4.0 International license
© Akiko Inoue, Tetsu Iwata, and Kazuhiko Minematsu

Integrity under the release of unverified plaintext (INT-RUP) is the security notion for authenticated encryption with associated data (AEAD) schemes. When an AEAD scheme is INT-RUP secure, it guarantees authenticity even when the decryption function inevitably or erroneously outputs the decrypted message without verification. INT-RUP security against existing AEAD schemes has been extensively studied, such as ChaCha20-Poly1305, SAEF, and TinyJambu. Many schemes have been designed with a provable security claim on INT-RUP as one of the main security features, such as Minalpher, CPF, LOTUS/LOCUS, and Oribatida. However, surprisingly, there is no INT-RUP analysis on GCM, which is one of the most widely deployed AEAD schemes. We prove that GCM has INT-RUP security with almost the same security level as that of the classical authenticity notion by showing that INT-RUP security of GCM is reduced to the variant of the unforgeability of GMAC inside GCM. Our future work is to analyze INT-RUP security on CCM.

3.13 Indifferentiability of 6-round Feistel

Mridul Nandi (Indian Statistical Institute - Kolkata, IN)

License © Creative Commons BY 4.0 International license
© Mridul Nandi
Joint work of Mridul Nandi, Ritam Bhaumik, Abishanka Saha, Paul Sayantan

The design and analysis of cryptographic systems often rely on proving the security of various primitives and constructions. One prominent framework for assessing the security of cryptographic constructions is indifferentiability introduced by Maurer et al. in [1], which provides a rigorous method to demonstrate the equivalence of constructed systems with idealized versions, even in the presence of adversaries with access to underlying primitives. This paper explores the indifferentiability of Feistel networks introduced by Horst Feistel as a design component of Lucifer [2], a widely used paradigm for constructing cryptographic primitives from simpler components known as round functions. Feistel networks offer a structured approach to building cryptographic systems, particularly permutations, by iteratively applying round functions to input data. Understanding the indifferentiability of Feistel networks is crucial for establishing the security of numerous cryptographic constructions.

A simulator to prove indifferentiability of six-round Feistel was proposed in [3] by Coron et al., along with an attack on five-round Feistel that works against any simulator. This simulator was later shown to be incapable of demonstrating indifferentiability of six-round Feistel by Holenstein et al. in [4]. An early version of [4] proposed a simulator for the 18-round construction that achieved indifferentiability, and in a later revision this was changed to a simulator for the 14-round version. Next, two concurrent works by Dai and Steinberger [5] and Dachman-Soled et al. [6] both proved indifferentiability of the 10-round Feistel network, the former by repairing the flawed 10-round simulator from [7], and the latter by modifying the 14-round simulator from [4]. Finally, in what has so far been the latest work in this series, Dai and Steinberger [8] established the indifferentiability of the 8-round Feistel construction

by optimising their simulator from [5]. Thus, while later proofs have been found for the indistinguishability of Feistel networks with eight or more rounds, that of six-round Feistel still remains an open question. The highlight of this paper is the presentation of a novel proof demonstrating the indistinguishability of the six-round Feistel network. This proof is accompanied by the design of an advanced simulator that preemptively handles construction queries from adversaries, ensuring the security of the constructed system.

References

- 1 U. Maurer, R. Renner, and C. Holenstein Indistinguishability, Impossibility Results on Reductions, and Applications to the Random Oracle Methodology. Theory of Cryptography Conference – TCC 2004, Lecture Notes in Computer Science, Springer-Verlag, vol. 2951, pp. 21–39, Feb 2004.
- 2 H. Feistel Cryptography and Computer Privacy *Scientific American* 228, no. 5 (1973): 15–23.
- 3 J. Coron, J. Patarin, Y. Seurin The Random Oracle Model and the Ideal Cipher Model Are Equivalent Advances in Cryptology – CRYPTO 2008. CRYPTO 2008. Lecture Notes in Computer Science, vol 5157. Springer, Berlin, Heidelberg.
- 4 T. Holenstein, R. Künzler, S. Tessaro Equivalence of the Random Oracle Model and the Ideal Cipher Model, Revisited. CoRR abs/1011.1264 (2010)
- 5 Y. Dai, J. Steinberger Indistinguishability of 10-Round Feistel Networks Eprint Paper 2015/874
- 6 D. Dachman-Soled, J. Katz, A. Thiruvengadam 10-Round Feistel is Indistinguishable from an Ideal Cipher EUROCRYPT (2) 2016: 649-678
- 7 Y. Seurin Primitives et protocoles cryptographiques à sécurité prouvée Université de Versailles Saint-Quentin-en-Yvelines
- 8 Y. Dai, J. Steinberger Indistinguishability of 8-Round Feistel Networks. CRYPTO (1) 2016: 95-120

3.14 The Algebraic Freelunch: Efficient Gröbner Basis Attacks Against Arithmetization-Oriented Primitives

Léo Perrin (INRIA – Paris, FR)

License © Creative Commons BY 4.0 International license
© Léo Perrin

Joint work of Augustin Bariant, Aurélien Boeuf, Axel Lemoine, Irati Manterola Ayala, Morten Øygarden, Léo Perrin, Håvard Raddum

Main reference Augustin Bariant, Aurélien Boeuf, Axel Lemoine, Irati Manterola Ayala, Morten Øygarden, Léo Perrin, Håvard Raddum: “The Algebraic Freelunch Efficient Gröbner Basis Attacks Against Arithmetization-Oriented Primitives”, IACR Cryptol. ePrint Arch., p. 347, 2024.

URL <https://eprint.iacr.org/2024/347>

In this talk, I presented a new type of algebraic attack that applies to many recent arithmetization-oriented families of permutations, such as those used in Griffin [1], Anemoi [2], and ArionHash [3], whose security relies on the hardness of the constrained-input constrained-output (CICO) problem. We introduce the FreeLunch approach: the monomial ordering is chosen so that the natural polynomial system encoding the CICO problem already is a Gröbner basis. In addition, we present a new dedicated resolution algorithm for FreeLunch systems of complexity lower than applicable state-of-the-art FGLM algorithms. We show that the FreeLunch approach challenges the security of fullround instances of Anemoi, Arion and Griffin. We confirm these theoretical results with experimental results on those three permutations. In particular, using the FreeLunch attack combined with a new technique to bypass 3 rounds of Griffin, we recover a CICO solution for 7 out of 10 rounds of Griffin in less than four hours on one core of AMD EPYC 7352 (2.3GHz).

References

- 1 Lorenzo Grassi, Yonglin Hao, Christian Rechberger, Markus Schofnegger, Roman Walch, Qingju Wang. *Horst meets fluid-SPN: Griffin for zero-knowledge applications*. Annual International Cryptology Conference. Cham: Springer Nature Switzerland, 2023.
- 2 Clémence Bouvier, Pierre Briaud, Pyrros Chaidos, Léo Perrin, Robin Salen, Vesselin Velichkov, Danny Willems. *New design techniques for efficient arithmetization-oriented hash functions: anemoi permutations and jive compression mode*. Annual International Cryptology Conference. Cham: Springer Nature Switzerland, 2023.
- 3 Arnab Roy, Matthias Johann Steiner, and Stefano Trevisani. *Arion: Arithmetization-oriented permutation and hashing from generalized triangular dynamical systems*. arXiv preprint arXiv:2303.04639 (2023).

3.15 The t-wise Independence of SPNs

Stefano Tessaro (University of Washington – Seattle, US)

License © Creative Commons BY 4.0 International license
© Stefano Tessaro

Joint work of Tianren Liu, Angelos Pelecanos, Stefano Tessaro, Vinod Vaikuntanathan

Main reference Tianren Liu, Angelos Pelecanos, Stefano Tessaro, Vinod Vaikuntanathan: “Layout Graphs, Random Walks and the t-Wise Independence of SPN Block Ciphers”, in Proc. of the Advances in Cryptology – CRYPTO 2023 – 43rd Annual International Cryptology Conference, CRYPTO 2023, Santa Barbara, CA, USA, August 20-24, 2023, Proceedings, Part III, Lecture Notes in Computer Science, Vol. 14083, pp. 694–726, Springer, 2023.


URL https://doi.org/10.1007/978-3-031-38548-3_23

This talk overviews an ongoing research agenda aimed at proving security of block ciphers against limited classes of attacks. We focus on proving that Substitution Permutation Networks (SPNs), when instantiated with concrete S-boxes (such as the AES S-box), give us t-wise independent permutations. This is a weak property (compared to being a full-fledged pseudorandom permutation), but it implies in particular security against classical families of statistical attacks such as linear and differential cryptanalysis. Our approach is in contrast with prior works aiming at full proofs of security for idealized versions of block ciphers.

4 Working groups

4.1 First results on the multivariate cryptanalysis of Poseidon

Lorenzo Grassi (Ruhr-Universität Bochum, DE), Antoine Joux (CISPA – Saarbrücken, DE), Patrick Neumann (Ruhr-Universität Bochum, DE), Léo Perrin (INRIA – Paris, FR), Christian Rechberger (TU Graz, AT), Ferdinand Sibleyras (NTT – Tokyo, JP), Aleksei Udovenko (University of Luxembourg, LU), and Qingju Wang (University of Luxembourg, LU)

License  Creative Commons BY 4.0 International license
 © Lorenzo Grassi, Antoine Joux, Patrick Neumann, Léo Perrin, Christian Rechberger, Ferdinand Sibleyras, Aleksei Udovenko, and Qingju Wang

The aim of this working group was to identify ways to improve algebraic attacks targeting arithmetization-oriented primitives (AOP) using “classical” tricks, i.e. insights gained not from a study of the system of equations, but from a careful analysis of the round function using well known approaches.

To this end, we first agreed to focus our efforts on Poseidon [1], one of the oldest AOPs. It is indeed an interesting target: it has a simple description, there is some public analysis of it, and yet its security against multi-variate approaches is ill-understood. In fact, at the time of the seminar, such analyses were (for Poseidon) essentially absent. Such a setting is made all the more relevant by recent advances in the zero-knowledge realm: the field sizes considered tend to be smaller nowadays, and the univariate approaches that were arguably the main threat at the time of the design of Poseidon have somewhat lost their relevance.

Due to the structure of the inner rounds of Poseidon, it is possible to bypass some of them when building a system of equations. In fact, the more words there are in the internal state, the more rounds we can by pass. Furthermore, since the round function of Poseidon has a low degree, we do not need to introduce an equation in each round. Thus, we believe that a CICO attack could work as follows.

1. introduce variables at the end of the first full rounds (so, after round 4);
2. write the equations encoding that the first input blocks are set to 0, which involves inverting the first rounds (while the round function is of low degree, its inverse is of very high degree, which is why we cover much fewer rounds backwards);
3. write the equations encoding that the first output blocks are set to 0, using the fact that the middle variables can be chosen so as to correspond to a subspace that does not activate the unique S-box in the middle rounds for a few rounds.
4. solve the equations using off-the-shelf tools.

It remains to precisely quantify the number of rounds that can be attacked in this fashion. We expect it to be lower than for a univariate attack, but, again, such attacks are not really relevant when the field size is smaller.

We also plan to make a detailed comparison between the case of Poseidon and that of Neptune, a very similar AOP that crucially differs in the structure of its outer rounds: instead of relying on low degree monomials (and thus, on functions with a very degree inverse), it relies on low degree functions that have the same degree in both directions. This should allow us to attack more rounds in this case, an insight we would consider valuable in itself.

References

- 1 Lorenzo Grassi, Dmitry Khovratovich, Christian Rechberger, Arnab Roy, Markus Schofnegger: *Poseidon: A new hash function for Zero-Knowledge proof systems*. 30th USENIX Security Symposium (USENIX Security 21). 2021.

4.2 Cryptanalysis of TEA3

Subhadeep Banik (University of Lugano, CH), Christof Beierle (Ruhr-Universität Bochum, DE), Anne Canteaut (INRIA – Paris, FR), Patrick Felke (Hochschule Emden/Leer, DE), Nils Gregor Leander (Ruhr-Universität Bochum, DE), Gaëtan Leurent (INRIA – Paris, FR), Yann Rotella (University of Versailles, FR), Sondre Rønjom (University of Bergen, NO), and Siwei Sun (University of Chinese Academy of Sciences, CN)

License © Creative Commons BY 4.0 International license
 © Subhadeep Banik, Christof Beierle, Anne Canteaut, Patrick Felke, Nils Gregor Leander, Gaëtan Leurent, Yann Rotella, Sondre Rønjom, and Siwei Sun

The proprietary TETRA Encryption standards, TEA1, TEA2, and TEA3, distributed by ETSI except for TEA2 (see [1]), were recently reverse-engineered in [2]. TEA1, TEA2, TEA3 are stream ciphers based on byte-oriented non-linear feedback shift registers (NFSRs). While TEA1 is an insecure cipher (the 80-bit key is compressed into a 4-byte register, effectively reducing its key length to 32 bits), the security level of the other algorithms TEA2 and TEA3 is less clear.

Although there is no obvious attack, the choices of components used in TEA3 (in contrast to TEA2) are questionable from a designer’s point of view. In particular, the key register of TEA3 employs an 8-bit Sbox in its feedback, denoted S , that is *not* a permutation. Instead, its distance to a permutation is only one bit in the sense that flipping a single bit in its lookup table would cause S to be bijective. Furthermore, the 10-byte key register can be decomposed into the cascade connection of a 5-byte NFSR into a 5-byte LFSR with feedback polynomial $(X^5 + SB3X^2 + 1) \cdot (X^5 + 1)$. Our experiments (started as joint work with Jens Alich, Christof Beierle, Patrick Felke, Gregor Leander, and Lukas Stennes) reveal that there are initial states of the 10-byte key register that have a period of only 10 bytes and that the maximal period is only about $10 \cdot 2^{40}$ bytes. The TEA3 key stream generator is depicted in [2, Figure 8]. The goal of this research group was to study the following questions:

- Can we exploit these properties to conduct an attack on TEA3?
- How were the components used in TEA3 designed?

During the research meetings at the Dagstuhl Seminar, the working group made the following progress on those questions:

- Using the existing theory on NFSRs from [3] and the structure of the cascade connection of the key register, we were able to derive necessary conditions on the period lengths of the key register. We further developed ideas to fully understand the cycle structure.
- The 16-to-8 bit functions $F31$ and $F32$ used in the state register of TEA3 are balanced vectorial Boolean functions built from 8 parallel 4-bit Boolean functions with overlapping inputs. A priori, it is not clear why such a construction leads to a balanced function. We identified a possible underlying construction method of $F31$ and $F32$.
- We identified some generic time-memory tradeoff attacks on the key stream generator with a complexity slightly below 2^{80} (i.e., the complexity of a brute-force attack)

References

- 1 ETSI. Custodian, security algorithms. <https://www.etsi.org/security-algorithms-and-codes/security-algorithms>, 2023. [Online; accessed 31-January-2024].
- 2 C. Meijer, W. Bokslag, and J. Wetzels. All cops are broadcasting: TETRA under scrutiny. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 7463–7479, 2023.
- 3 J. Mykkeltveit, M.-K. Siu, and P. Tong. On the cycle structure of some nonlinear shift register sequences. *Information and control*, 43(2):202–215, 1979.

4.3 Exploitation of the Wrong Key Randomization Hypothesis Non-conformity in Key Recovery Attacks

Zhenzhen Bao (*Tsinghua University – Beijing, CN*) and Nils Gregor Leander (*Ruhr-Universität Bochum, DE*)

License © Creative Commons BY 4.0 International license
© Zhenzhen Bao and Nils Gregor Leander

Main reference Aron Gohr: “Improving Attacks on Round-Reduced Speck32/64 Using Deep Learning”, in Proc. of the Advances in Cryptology – CRYPTO 2019 – 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part II, Lecture Notes in Computer Science, Vol. 11693, pp. 150–179, Springer, 2019.

URL https://doi.org/10.1007/978-3-030-26951-7_6

At CRYPTO 2019, Gohr introduced a novel key recovery strategy for Speck32/64, termed the BayesianKeySearch algorithm. This approach challenges the conventional Wrong Key Randomization Hypothesis (WKRH), particularly in scenarios where the attack on Speck undergoes only a single round of trial decryption. The BayesianKeySearch algorithm iteratively refines key guesses by generating new round-key candidates based on previous guesses, thereby progressively enhancing the quality of the guessed key. Notably, this method requires a limited number of one-round trial decryptions, reducing time complexity.

Despite its practical implications, the algorithm lacks a comprehensive theoretical framework to quantify the impact of various parameters on attack complexity and success rate. This discussion group aimed to bridge this gap by establishing a theoretical model to evaluate its complexity better and broadening the application to conventional key recovery attacks.

The discussion centered on two primary aspects: examining how the wrong key-right key distance influences the statistics in traditional differential and linear attacks and developing methods to leverage these non-random influences. Initial outcomes of this discussion include a proposed formula hypothesizing the relationship between the Hamming weight of the wrong key-right key distance and the probability of returning to the same output difference, a drawn parallel between this probability’s evaluation and the BCT’s computation (difference-boomerang connective probability), and a proposed time-data tradeoff strategy in differential key-recovery attacks.

The group plans to persist in the investigation of this topic post-seminar.

4.4 Cryptanalysis of SCARF

Christina Boura (*University of Versailles, FR*), Zahra Ahmadian (*Shahid Beheshti University – Tehran, IR*), Yanis Belkheyar (*Radboud University Nijmegen, NL*), Christoph Dobraunig (*Intel – Villach, AT*), Henri Gilbert (*ANSSI – Paris, FR*), Shahram Rasoolzadeh (*Radboud University Nijmegen, NL*), Dhiman Saha (*Indian Institute of Technology Bhilai – Durg, IN*), Tyge Tiessen (*Technical University of Denmark – Lyngby, DK*), and Yosuke Todo (*NTT – Tokyo, JP*)

License © Creative Commons BY 4.0 International license

© Christina Boura, Zahra Ahmadian, Yanis Belkheyar, Christoph Dobraunig, Henri Gilbert, Shahram Rasoolzadeh, Dhiman Saha, Tyge Tiessen, and Yosuke Todo

SCARF if a tweakable BC for cache randomization, proposed at Usenix Security 2023.

It’s block length is 10 bits only, it absorbs a tweak of 48 bits and it offers 80-bit security (but a 240-bit key is used). The security claim is peculiar as an attacker cannot directly query the TBC. Instead, the attacker can query collision or composition oracles. The query complexity is up to 2^{40} .

In this working group we analyzed the security of SCARF by investigating different cryptanalysis techniques. We thought of several approaches:

- Exploit the fact that 0 is a fixed point for some operations of the round function.
- Analyse the algebraic degree growth and exploit the algebraic normal form of the Sbox.
- Polytopic cryptanalysis.
- Multiple-tweak differential attack.

While all the approaches seem promising, the multiple-tweak differential attack seems to be particularly interesting for this cipher and we hope to be able to break $(7+7)/(8+8)$ rounds of this cipher with this approach.

References

- 1 Canale F., Güneysu T., Leander G., Thoma J. P., Todo Y., Ueno R. *A Low-Latency Block Cipher for Secure Cache-Randomization*. USENIX Security Symposium 2023: 1937-1954

4.5 Differential cryptanalysis and more

Patrick Derbez (University of Rennes, FR), Orr Dunkelman (University of Haifa, IL), Maria Eichlseder (TU Graz, AT), Ryoma Ito (NICT – Tokyo, JP), Virginie Lallemand (LORIA – Nancy, FR), and María Naya-Plasencia (INRIA – Paris, FR)

License © Creative Commons BY 4.0 International license
© Patrick Derbez, Orr Dunkelman, Maria Eichlseder, Ryoma Ito, Virginie Lallemand, and María Naya-Plasencia

Our research group was composed of Maria Eichlseder, Orr Dunkelman, María Naya-Plasencia, Virginie Lallemand, Ryoma Ito and Patrick Derbez. The main topic of our group was to propose a new modelization of impossible differential attacks, more accurate than a classic “0/1/?” model but faster than exhausting all differential characteristics. Our idea is to track as well the equalities between internal state variables to propagate more information and extend the impossible cases. We also investigated several other topics including extension of differential-mitm attacks, impossible differential-linear attacks as well as an attack against a generic cipher relying on the nice algorithmic problem of finding the closest pair of vectors.

4.6 Key Control Security Group

Tetsu Iwata (Nagoya University, JP), Ritam Bhaumik (EPFL – Lausanne, CH), Avijit Dutta (TCG CREST – Kolkata, IN), Akiko Inoue (NEC – Kawasaki, JP), Ashwin Jha (Ruhr-Universität Bochum, DE), Kazuhiko Minematsu (NEC – Kawasaki, JP), Mridul Nandi (Indian Statistical Institute – Kolkata, IN), Yu Sasaki (NTT – Tokyo, JP), Meltem Sonmez Turan (NIST – Gaithersburg, US), Stefano Tessaro (University of Washington – Seattle, US), and Aishwarya Thiruvengadam (Indian Institute of Technology Madras, IN)

License © Creative Commons BY 4.0 International license
© Tetsu Iwata, Ritam Bhaumik, Avijit Dutta, Akiko Inoue, Ashwin Jha, Kazuhiko Minematsu, Mridul Nandi, Yu Sasaki, Meltem Sonmez Turan, Stefano Tessaro, and Aishwarya Thiruvengadam

NIST SP 800-108r1 [1] specifies Key Derivation Functions (KDFs) based on PseudoRandom Functions (PRFs). It specifies a KDF based on KMAC, and it also specifies KDFs in counter mode, feedback mode, and double-pipeline mode, which are combined with HMAC or CMAC as the PRF.

The document was revised in August 2022, and a discussion on the key control security was added, showing a security issue in KDFs with CMAC. The goal of this research group is to formalize a cryptographic definition of the key control security, and analyze the security of KDFs in the NIST document from the provable security and cryptanalytic perspectives.

The group identified a formal security definition, and drafted a security proof of a KDF based on KMAC. We also analyzed KDFs based on CMAC from a cryptanalytic view point. In particular, we focused on the strengthened version, which is a variant of the KDFs based on CMAC to mitigate the issue in their key control security.

References

- 1 Lily Chen. Recommendation for Key Derivation Using Pseudorandom Functions. NIST Special Publication, NIST SP 800-108r1, August 2022, <https://doi.org/10.6028/NIST.SP.800-108r1>

4.7 Security of sponge combiners

Charlotte Lefevre (Radboud University Nijmegen, NL), Rachelle Heim Boissier (University of Versailles, FR), Bart Mennink (Radboud University Nijmegen, NL), and Bart Preneel (KU Leuven, BE)

License © Creative Commons BY 4.0 International license
© Charlotte Lefevre, Rachelle Heim Boissier, Bart Mennink, and Bart Preneel

The aim of this research group was to investigate the security of sponge-based combiners and variants from both cryptanalytical and provable security perspectives. The motivation behind this investigation stemmed from the observation that while finding inner collisions in a sponge generically requires $2^{c/2}$ permutation evaluations, the associated attack does not straightforwardly generalize with combiners due to the repeated absorption of the same message block. On the cryptanalytical aspect, we examined various sponge-based combiners and derivatives, namely the concatenation combiner, XOR combiner, and two hash-twice-like constructions. We focused on collision, second preimage, and preimage attacks, with Joux's attack [1] serving as the main tool. Notably, except for collision of the hash-twice construction with identical permutations (which costs $2^{c/2}$ evaluations), a term in $2^{b/2}$ emerged consistently in our analysis. On the provable security aspect, the discussions provided valuable insights that could be helpful to improve the security of these constructions, ideally up to $\min(c, b/2)$ bits. We plan to continue working on these two aspects after the seminar.

References

- 1 A. Joux. Multicollisions in Iterated Hash Functions. *Advances in Cryptology – Crypto 2004*, Volume 3152 of Lecture Notes in Computer Science. Springer-Verlag, 2004.

Participants

- Zahra Ahmadian
Shahid Beheshti University –
Tehran, IR
- Subhadeep Banik
University of Lugano, CH
- Zhenzhen Bao
Tsinghua University –
Beijing, CN
- Christof Beierle
Ruhr-Universität Bochum, DE
- Yanis Belkheyar
Radboud University
Nijmegen, NL
- Ritam Bhaumik
EPFL – Lausanne, CH
- Christina Boura
University of Versailles, FR
- Anne Canteaut
INRIA – Paris, FR
- Patrick Derbez
University of Rennes, FR
- Christoph Dobraunig
Intel – Villach, AT
- Orr Dunkelman
University of Haifa, IL
- Avijit Dutta
TCG CREST – Kolkata, IN
- Maria Eichlseder
TU Graz, AT
- Patrick Felke
Hochschule Emden/Leer, DE
- Henri Gilbert
ANSSI – Paris, FR
- Lorenzo Grassi
Ruhr-Universität Bochum, DE
- Rachelle Heim Boissier
University of Versailles, FR
- Akiko Inoue
NEC – Kawasaki, JP
- Ryoma Ito
NICT – Tokyo, JP
- Tetsu Iwata
Nagoya University, JP
- Ashwin Jha
Ruhr-Universität Bochum, DE
- Antoine Joux
CISPA – Saarbrücken, DE
- Virginie Lallemand
LORIA – Nancy, FR
- Nils Gregor Leander
Ruhr-Universität Bochum, DE
- Charlotte Lefevre
Radboud University
Nijmegen, NL
- Gaëtan Leurent
INRIA – Paris, FR
- Willi Meier
FH Nordwestschweiz –
Windisch, CH
- Bart Mennink
Radboud University
Nijmegen, NL
- Kazuhiko Minematsu
NEC – Kawasaki, JP
- Mridul Nandi
Indian Statistical Institute –
Kolkata, IN
- María Naya-Plasencia
INRIA – Paris, FR
- Patrick Neumann
Ruhr-Universität Bochum, DE
- Léo Perrin
INRIA – Paris, FR
- Bart Preneel
KU Leuven, BE
- Shahram Rasoolzadeh
Radboud University
Nijmegen, NL
- Christian Rechberger
TU Graz, AT
- Yann Rotella
University of Versailles, FR
- Sondre Rønjom
University of Bergen, NO
- Dhiman Saha
Indian Institute of Technology
Bhilai – Durg, IN
- Yu Sasaki
NTT – Tokyo, JP
- Ferdinand Sibleyras
NTT – Tokyo, JP
- Meltem Sonmez Turan
NIST – Gaithersburg, US
- Siwei Sun
University of Chinese Academy
of Sciences, CN
- Stefano Tessaro
University of Washington –
Seattle, US
- Aishwarya Thiruvengadam
Indian Institute of Technology
Madras, IN
- Tyge Tiessen
Technical University of Denmark
– Lyngby, DK
- Yosuke Todo
NTT – Tokyo, JP
- Aleksei Udovenko
University of Luxembourg, LU
- Qingju Wang
University of Luxembourg, LU



The Emerging Issues in Bioimaging AI Publications and Research

Jianxu Chen^{*1}, Florian Jug^{*2}, Susanne Rafelski^{*3}, and Shanghang Zhang^{*4}

- 1 ISAS – Dortmund, DE. jianxu.chen@isas.de
- 2 Human Technopole – Milano, IT. florian.jug@fht.org
- 3 Allen Institute for Cell Science – Seattle, US. susanner@alleninstitute.org
- 4 Peking University, CN. shanghang@pku.edu.cn

Abstract

This report documents the program and outcomes of Dagstuhl Seminar “The Emerging Issues in Bioimaging AI Publications and Research” (24042) held on January 21–24, 2024. The fast advancement of computational techniques, particularly those based on artificial intelligence (AI), has significantly propelled the field of computational biology. With the rapid development, new issues are emerging in bioimaging AI publications and research. For example, how can we properly validate the AI methods used in quantitative biological analysis? Also, the ethical aspects of these developments remain underexplored, lacking clear definitions and recognition within the community. The goal of this interdisciplinary seminar was to bring together experts from various fields, including experimental biology, computational biology, bioimage analysis, computer vision, and AI research, to identify, discuss and address the emerging issues in current bioimaging AI research and publications.

Seminar January 21–24, 2024 – <https://www.dagstuhl.de/24042>

2012 ACM Subject Classification Applied computing → Imaging; Computing methodologies → Artificial intelligence

Keywords and phrases artificial intelligence, bioimaging, open source, publication ethics, trustworthy ai

Digital Object Identifier 10.4230/DagRep.14.1.90


1 Executive Summary

Jianxu Chen (ISAS – Dortmund, DE)

Florian Jug (Human Technopole – Milano, IT)

Susanne Rafelski (Allen Institute for Cell Science – Seattle, US)

Shanghang Zhang (Peking University, CN)

License  Creative Commons BY 4.0 International license
© Jianxu Chen, Florian Jug, Susanne Rafelski, and Shanghang Zhang

Seminar Structure and Organization

The seminar was divided into three specific directions: ethical considerations in bioimaging AI research and publications, performance reporting on bioimaging AI methods in publications and research, and future research directions of bioimaging AI focusing on validation and robustness. The seminar was structured into two parts: the first half focused on presentations and information sharing related to these three major directions to align experts from different fields, and the second half concentrated on in-depth discussions of these topics.

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

The Emerging Issues in Bioimaging AI Publications and Research, *Dagstuhl Reports*, Vol. 14, Issue 1, pp. 90–107
Editors: Jianxu Chen, Florian Jug, Susanne Rafelski, and Shanghang Zhang

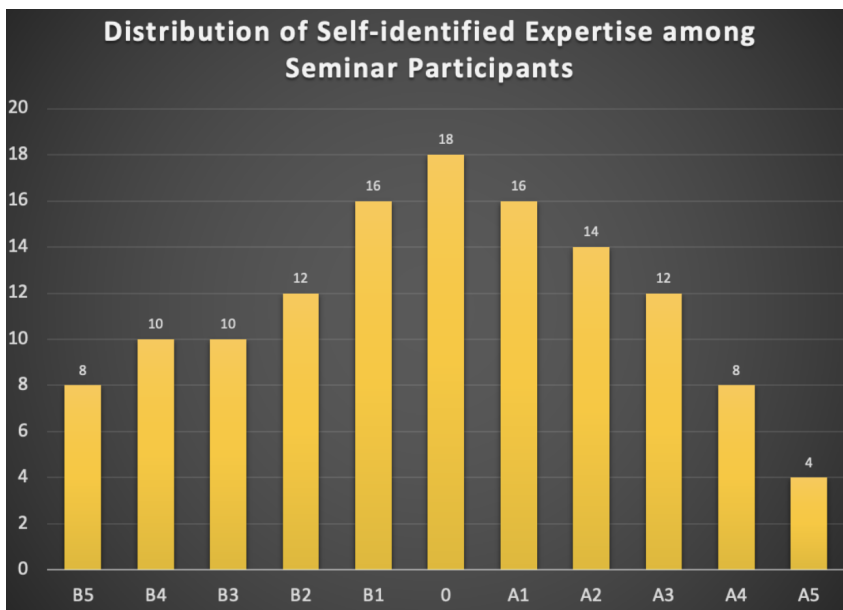


Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Given the highly interdisciplinary nature of the seminar, we took two specific steps to facilitate smooth communication and discussion among researchers with diverse backgrounds.

First, about six to eight weeks before the seminar, we sent out a survey to gather potential topics each participant could present within the seminar’s overarching theme. We collaborated with several participants to choose or adjust their presentation topics to ensure the effectiveness in this interdisciplinary setting. Based on the survey responses, the presentation and information-sharing portion (the first half) of our seminar began with two keynotes from editors who handle bioimaging AI papers, sharing their insights and the existing efforts by publishers. We then organized all presentations to progress from a focus on biology to bioimaging AI, and finally to AI, ensuring coverage of the full spectrum of necessary knowledge for our in-depth discussions in the second half.

Second, at the beginning of the seminar, we allocated two minutes for each participant for a quick introduction and to briefly rate their experience and expertise on a scale in the range of [B5, B4, B3, B2, B1, 0, A1, A2, A3, A4, A5], with B5 representing pure biology and A5 representing pure AI. Participants could select a single value, multiple values, or a range of values. This was not intended to stereotype participants but to facilitate easier communication. For example, if a participant with experience in the range of B5 to B3 spoke with two others during a coffee break, one with experience from B3 to A1 and the other from A3 to A5, different communication strategies would be necessary for effective discussions. The distribution of self-identified experience is summarized in the histogram below (see Fig. 1).



■ **Figure 1** Histogram of the distribution of expertised self-identified by seminar participants.

Presentations, discussions and outcomes

Overview of the scientific talks

The seminar began with presentations by editors from Nature Methods and Cell Press, who shared their insights on existing and emerging issues in bioimaging AI publications. Following this, general bioimage analysis validation issues were discussed from both a biological application perspective and an algorithmic metric perspective. These presentations were succeeded by specific application talks demonstrating how AI-based bioimage analysis is utilized and validated in high-throughput biological applications [1]. The remainder of day one focused on bioimaging AI validation through explainable AI [2], [3], [4] and existing tools [5], as well as community efforts in deploying FAIR (Findable, Accessible, Interoperable, Reusable) AI tools for bioimage analysis [6].

The second day commenced with several theoretical AI talks introducing key concepts related to model robustness, fairness, and trustworthiness [7]. These were followed by two presentations showcasing state-of-the-art AI algorithms applied in bioimaging [8], [9], and an overview of the application of foundation models in bioimaging [10]. The scientific presentation portion of the seminar concluded with a talk about the pilot work initiated by the EMBO (European Molecular Biology Organization) Press on research integrity and AI integration in publishing and trust. This talk also served as a transition into the in-depth discussions that comprised the second part of the seminar.

Summary of discussions and key outcomes

After the scientific presentation part of the seminar, the participants naturally reach the agreement on doing the discussion in a four-quadrant manner, as illustrated below in Fig. 2.

	“Users” of bioimaging AI	“Makers” of bioimaging AI
In-domain technical considerations	1	2
out-of-domain implications	3	4

■ **Figure 2** The four-quadrant for organizing the in-depth discussion.

Here are some examples of what emerges from discussions in each quadrant.

I. What are some technical considerations that users of AI should pay attention to?

When using a specific bioimage analysis model, it is crucial for users to have clear biological questions that align with the technical limitations of the bioimaging AI models. This is known as application-appropriate validation [11]. For example, the trustworthiness or validity of an AI-based microscopy image denoising model may differ significantly between a study that requires merely counting the number of nuclei in an image and one that aims to quantify the morphological properties of the nuclei.

II. What are some technical considerations that makers of AI should pay attention to?

When developing a bioimaging AI model, comprehensive evaluations and ablation studies are essential to explicitly demonstrate the model’s limitations or potential failures. For instance, evaluating a cell segmentation model under different conditions, such as various magnifications, signal-to-noise ratios, cell densities, and possibly different microscope modalities, is highly

among participants, including wet-lab biologists and machine learning theorists with little biology experience, created unique networking opportunities. They would otherwise have rare opportunities to meet in traditional conferences. Biologists expressed that they gained new insights into the theories behind machine learning methods they had used, motivating them to rethink their future research designs. Conversely, machine learning researchers showed strong interest in collaborating with the bioimaging community to address fundamental challenges such as robustness and explainability.

Conclusions

This Dagstuhl Seminar on “The Emerging Issues in Bioimaging AI Publications and Research” successfully united a diverse group of experts from experimental biology, computational biology, bioimage analysis, computer vision, and AI research. The seminar facilitated in-depth discussions on ethical considerations, performance reporting, and future research directions in bioimaging AI, highlighting the crucial need for interdisciplinary collaboration and communication.

Through structured presentations and interactive discussions, participants underscored the importance of clear communication between AI developers and users, comprehensive model validation, and awareness of biological batch effects. The seminar emphasized the necessity for application-appropriate validation and detailed reporting of AI model conditions to enhance the trustworthiness and applicability of bioimaging AI methods. Furthermore, the seminar provided a valuable platform for social interactions and networking, bridging gaps between researchers from different fields and fostering new collaborations.

In conclusion, the seminar not only advanced discussions on critical issues in bioimaging AI publications but also laid the foundation for ongoing collaboration and innovation in the field. Planned follow-up activities will further contribute to the development and ethical application of AI in bioimaging research. The success of this seminar underscores the importance of continuous communication and cooperation in addressing the emerging challenges in bioimaging AI publications and research.

Acknowledgement

We are grateful to all seminar participants for their insightful contributions and the engaging discussions they fostered, especially in the interdisciplinary setting with a wide spectrum of expertise. We also sincerely thank the Dagstuhl Scientific Directorate for the opportunity to organize this event. Finally, our deepest appreciation goes to the exceptional Dagstuhl staff whose support was instrumental in making the seminar a success.

References

- 1 Z. Cibir et al., “ComplexEye: a multi-lens array microscope for high-throughput embedded immune cell migration analysis,” *Nat. Commun.*, vol. 14, no. 1, p. 8103, Dec. 2023, doi: 10.1038/s41467-023-43765-3.
- 2 Christopher J. Soelistyo and Alan R. Lowe, “Discovering Interpretable Models of Scientific Image Data with Deep Learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2024*, pp. 6884–6893. [Online].
- 3 Christopher J. Soelistyo, Guillaume Charras, and Alan R. Lowe, “Virtual Perturbations to Assess Explainability of Deep-Learning Based Cell Fate Predictors,” in *Proceedings of the*

- IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 2023, pp. 3971–3980. [Online].
- 4 D. Schuhmacher et al., “A framework for falsifiable explanations of machine learning models with an application in computational pathology,” *Med. Image Anal.*, vol. 82, p. 102594, Nov. 2022, doi: 10.1016/j.media.2022.102594.
 - 5 L. M. Moser et al., “Piximi – An Images to Discovery web tool for bioimages and beyond.” Jun. 04, 2024. doi: 10.1101/2024.06.03.597232.
 - 6 W. Ouyang et al., “BioImage Model Zoo: A Community-Driven Resource for Accessible Deep Learning in BioImage Analysis,” *Bioinformatics*, preprint, Jun. 2022. doi: 10.1101/2022.06.07.495102.
 - 7 D. Guo, C. Wang, B. Wang, and H. Zha, “Learning Fair Representations via Distance Correlation Minimization,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 2139–2152, Feb. 2024, doi: 10.1109/TNNLS.2022.3187165.
 - 8 Saumya Gupta, Yikai Zhang, Xiaoling Hu, and Prateek Prasanna, “Topology-aware uncertainty for image segmentation,” presented at the Advances in Neural Information Processing Systems, 2024.
 - 9 G. Dai et al., “Implicit Neural Image Field for Biological Microscopy Image Compression.” arXiv, 2024. doi: 10.48550/ARXIV.2405.19012.
 - 10 A. Archit et al., “Segment Anything for Microscopy,” *Bioinformatics*, preprint, Aug. 2023. doi: 10.1101/2023.08.21.554208.
 - 11 J. Chen, M. P. Viana, and S. M. Rafelski, “When seeing is not believing: application-appropriate validation matters for quantitative bioimage analysis,” *Nat. Methods*, vol. 20, no. 7, pp. 968–970, Jul. 2023, doi: 10.1038/s41592-023-01881-4.
 - 12 J. Chen et al., “The Allen Cell and Structure Segmenter: a new open source toolkit for segmenting 3D intracellular structures in fluorescence microscopy images,” *Cell Biology*, preprint, Dec. 2018. doi: 10.1101/491035.

2 Table of Contents

Executive Summary

Jianxu Chen, Florian Jug, Susanne Rafelski, and Shanghang Zhang 90

Overview of Talks

Topological Uncertainty and Representation in Biomedical Image Analysis
Chao Chen 97

Application-appropriate validation matters for quantitative bioimage analysis
Jianxu Chen, Matheus Palhares Viana, and Susanne Rafelski 98

Metrics reloaded: Recommendations for image analysis validation
Evangelia Christodoulou 98

Working towards pick 5: strategies for scaling and distributing user-friendly containers
Beth Cimini 99

Implicit Neural Representation (INR) for Biological Image Compression and Neural Plasticity Learning
Gaole Dai 100

An overview of Cell Press policies on image presentation, data and code sharing, and AI use
Andrew Hufton 100

Discovering interpretable models of scientific image data with deep learning
Alan Lowe 101

Improving trustworthiness of ML in bioimaging through experimentally testable explanations
Axel Mosig 101

Segment Anything for Microscopy
Constantin Pape 103

High-volume, label-free imaging for quantifying single-cell dynamics in iPSC colonies
Anne Plant 103

Publishing microscopy and AI in Nature Methods
Rita Strack 104

Frequency shortcuts learning and generalization in computer vision
Nicola Strisciuglio 104

Visual interpretability of deep learning models in cell imaging
Assaf Zaritsky 105

Foundation models for biomedical image analysis
Shanghang Zhang 105

Algorithmic Fairness, Robust Generalization and Trustworthy Machine Learning
Han Zhao 106

Participants 107

3 Overview of Talks

3.1 Topological Uncertainty and Representation in Biomedical Image Analysis

Chao Chen (Stony Brook University, US)

- License** © Creative Commons BY 4.0 International license
© Chao Chen
- Main reference** Xiaoling Hu, Fuxin Li, Dimitris Samaras, Chao Chen: “Topology-Preserving Deep Image Segmentation”, in Proc. of the Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, pp. 5658–5669, 2019.
URL <https://proceedings.neurips.cc/paper/2019/hash/2d95666e2649fcfc6e3af75e09f5adb9-Abstract.html>
- Main reference** Saumya Gupta, Yikai Zhang, Xiaoling Hu, Prateek Prasanna, Chao Chen: “Topology-Aware Uncertainty for Image Segmentation”, 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/19ded4cfc36a7feb7fce975393d378fd-Abstract-Conference.html
- Main reference** Shahira Abousamra, Rajarsi Gupta, Tahsin M. Kurç, Dimitris Samaras, Joel H. Saltz, Chao Chen: “Topology-Guided Multi-Class Cell Context Generation for Digital Pathology”, in Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023, pp. 3323–3333, IEEE, 2023.
URL <https://doi.org/10.1109/CVPR52729.2023.00324>

Modern analytics is facing highly complex and heterogeneous data. While deep learning models have pushed our prediction power to a new level, they are not satisfactory in some crucial merits such as transparency, robustness, data-efficiency, etc. To address these challenges, I am generally interested in incorporating mathematical modeling of topology, geometry and dynamics seamlessly into the learning pipeline. Such model-informed learning approach will be more transparent, steerable and less annotation-hungry.

In this talk, I will focus on our recent work on combining topological reasoning with learning to solve problems in biomedical image analysis. With advanced imaging techniques, we are collecting images of various complex structures such as neurons, vessels, tissues and cells. These structures encode important information about underlying biological mechanisms. To fully exploit these structures, we propose to enhance learning pipelines with topology, the branch of abstract mathematics that deals with structures such as connections, loops and branches. Under the hood is a formulation of the topological computation as a robust and differentiable operator, based on the theory of persistent homology. This inspires a series of novel methods for segmentation, uncertainty estimation, generation, and analysis of these topology-rich biomedical structures.

3.2 Application-appropriate validation matters for quantitative bioimage analysis

Jianxu Chen (ISAS – Dortmund, DE), Matheus Palhares Viana (Allen Institute for Cell Science – Seattle, US), and Susanne Rafelski (Allen Institute for Cell Science – Seattle, US)

License © Creative Commons BY 4.0 International license

© Jianxu Chen, Matheus Palhares Viana, and Susanne Rafelski

Main reference Jianxu Chen, Matheus P. Viana, and Susanne Rafelski: “When seeing is not believing: application-appropriate validation matters for quantitative bioimage analysis”. *Nat Methods* 20, 968–970 (2023).

URL <https://doi.org/10.1038/s41592-023-01881-4>

A critical step towards biologically reliable analysis of microscopy image-based assays is rigorous quantitative validation with metrics and measurements appropriate for the particular biological application. Currently, however, no community standards or publication guidelines exist on how to conduct the appropriate validation of work involving quantitative analysis of microscopy images, including deep-learning based approaches. In this presentation, we discussed this challenge for both classical and modern deep-learning based image analysis approaches as well as possible solutions for automating and streamlining the validation process. First, to introduce the concept of “application-appropriate validation”, we showed a true story of how inappropriate validation of segmentations in a quantitative analysis of mitochondrial network morphology led to wrong biological conclusions. Second, besides segmentation, we showed another example of application-appropriate validation for label-free predictions. The commonly used metrics, e.g., Pearson correlation or structure similarity, could be misleading when not taking the downstream biological application into account. Finally, we discussed a list of key considerations for interpretable quantification of microscopy image-based assays, from understanding the underlying biological questions, understanding the limits of assays, understanding the validation requirement for interpretation, to the estimation of time and effort one could afford, with an emphasis on future community efforts in standardization, dissemination and interdisciplinary connections.

3.3 Metrics reloaded: Recommendations for image analysis validation

Evangelia Christodoulou (DKFZ – Heidelberg, DE)

License © Creative Commons BY 4.0 International license

© Evangelia Christodoulou

Main reference Lena Maier-Hein, Annika Reinke, Evangelia Christodoulou, Ben Glocker, Patrick Godau, Fabian Isensee, Jens Kleesiek, Michal Kozubek, Mauricio Reyes, Michael A. Riegler, Manuel Wiesenfarth, Michael Baumgartner, Matthias Eisenmann, Doreen Heckmann-Nötzl, A. Emre Kavur, Tim Rüdtsch, Minu Dietlinde Tizabi, Laura Ación, Michela Antonelli, Tal Arbel, Spyridon Bakas, Peter Bankhead, Arriel Benis, M. Jorge Cardoso, Veronika Cheplygina, Beth A. Cimini, Gary S. Collins, Keyvan Farahani, Bram van Ginneken, Daniel A. Hashimoto, Michael M. Hoffman, Merel Huisman, Pierre Jannin, Charles E. Kahn, Alexandros Karargyris, Alan Karthikesalingam, Hannes Kenngott, Annette Kopp-Schneider, Anna Kreshuk, Tahsin M. Kurç, Bennett A. Landman, Geert Litjens, Amin Madani, Klaus H. Maier-Hein, Anne L. Martel, Peter Mattson, Erik Meijering, Bjoern H. Menze, David Moher, Karel G. M. Moons, Henning Müller, Felix Nickel, Brennan Nichyporuk, Jens Petersen, Nasir M. Rajpoot, Nicola Rieke, Julio Saez-Rodriguez, Clarisa Sánchez Gutiérrez, Shravya Shetty, Maarten van Smeden, Carole H. Sudre, Ronald M. Summers, Abdel A. Taha, Sotirios A. Tsaftaris, Ben Van Calster, Gaël Varoquaux, Paul F. Jäger: “Metrics reloaded: Pitfalls and recommendations for image analysis validation”, *CoRR*, Vol. abs/2206.01653, 2022.

URL <https://doi.org/10.48550/ARXIV.2206.01653>

Increasing evidence shows that flaws in machine learning (ML) algorithm validation are an underestimated global problem. Particularly in automatic biomedical image analysis, chosen performance metrics often do not reflect the domain interest, thus failing to adequately

measure scientific progress and hindering translation of ML techniques into practice. To overcome this, our large international expert consortium created Metrics Reloaded, a comprehensive framework guiding researchers in the problem-aware selection of metrics. Following the convergence of ML methodology across application domains, Metrics Reloaded fosters the convergence of validation methodology. The framework was developed in a multi-stage Delphi process and is based on the novel concept of a problem fingerprint – a structured representation of the given problem that captures all aspects that are relevant for metric selection, from the domain interest to the properties of the target structure(s), data set and algorithm output. Based on the problem fingerprint, users are guided through the process of choosing and applying appropriate validation metrics while being made aware of potential pitfalls. Metrics Reloaded targets image analysis problems that can be interpreted as a classification task at image, object or pixel level, namely image-level classification, object detection, semantic segmentation, and instance segmentation tasks. To improve the user experience, we implemented the framework in the Metrics Reloaded online tool, which also provides a point of access to explore weaknesses, strengths and specific recommendations for the most common validation metrics. The broad applicability of our framework across domains is demonstrated by an instantiation for various biological and medical image analysis use cases.

3.4 Working towards pick 5: strategies for scaling and distributing user-friendly containers

Beth Cimini (Broad Institute of MIT & Harvard – Cambridge, US)

License  Creative Commons BY 4.0 International license
© Beth Cimini

In the current scientific software environment, we have identified 5 axes that should be measured for any software or code distribution system with which one plans to share code.

Reproducible – can you tell what went in there and why and how?

Easy to create – how much extra work/knowledge is needed on the developer side to package?

Easy to run – how much extra work/knowledge is needed on the user side to run?


Long lasting – will the thing I made today work tomorrow?

Scalable – if my experiments get bigger (in terms of individual image size and/or parallelization of many images, can I still use my solution?

In this talk, we discuss the merits of packaged applications, virtual environment spec files, online workflow tools like Galaxy, as well as software containers. We discuss strategies that can be used alongside software containers to make them maximally user friendly, and discuss possible strategies to make containers score high on all 5 axes.

3.5 Implicit Neural Representation (INR) for Biological Image Compression and Neural Plasticity Learning

Gaole Dai (*Peking University, CN*)

License  Creative Commons BY 4.0 International license
© Gaole Dai

The presentation introduces Implicit Neural Representation (INR), which is employed for various tasks such as image compression and 3D reconstruction. Specifically, we explore the distinctive attributes of INR in bioimage data compression. Additionally, we investigate the integration of the coordinate-to-value learning approach of INR into conventional Artificial Neural Networks (ANNs). By assigning specific coordinates to each cell/synapse in the ANN and integrating them with the INR network, we obtain tailored adjustment values for each location. We find that this type of adjustment exhibits neural plasticity, a characteristic unique to biological networks, making it highly valuable in Parameter Efficient Fine-Tuning (PEFT) tasks.

3.6 An overview of Cell Press policies on image presentation, data and code sharing, and AI use

Andrew Hufton (*Patterns, Cell Press – Würzburg, DE*)

License  Creative Commons BY 4.0 International license
© Andrew Hufton

The Cell Press journals, including *Patterns* (<https://www.cell.com/patterns/>), have high standards for the transparency and reproducibility of research presented at our journals. In my talk, I presented a brief overview of our policies on image presentation, data and code sharing, and the use of AI tools in research and manuscript preparation. I then discussed how these policies apply to cutting-edge bioimaging research and some of the challenges editors and our authors commonly face during the peer-review and publication process. Notably, I made the case that authors should think critically about the openness, ethics and transparency of AI models and training datasets used in their research, and should keep in mind that reliance on closed-source commercial models could impact the transparency and publishability of their work. I also highlighted some of the dangers of poorly-designed AI detection tools, and argued that while we must be vigilant against AI-enabled fraud, our main focus as a community should be on positively promoting and rewarding innovative, rigorous and open research. A selection of papers mentioned in my talk are included below.

References

- 1 Bagheri, N., et al (2023) The new era of quantitative cell imaging—challenges and opportunities. *Mol. Cell* 82, 241-247. <https://doi.org/10.1016/j.molcel.2021.12.024>
- 2 Gu, J., et al (2022) AI-enabled image fraud in scientific publications. *Patterns* 3, 100511. <https://doi.org/10.1016/j.patter.2022.100511>
- 3 Liang, W., et al (2023) GPT detectors are biased against non-native English writers. *Patterns* 4, 100779. <https://doi.org/10.1016/j.patter.2023.100779>
- 4 Wang, W., et al. (2023) On the transparency of large AI models. *Patterns* 4, 100797. <https://doi.org/10.1016/j.patter.2023.100797>

3.7 Discovering interpretable models of scientific image data with deep learning

Alan Lowe (*The Alan Turing Institute – London, GB*)

License © Creative Commons BY 4.0 International license
© Alan Lowe

Joint work of Christopher Soelistyo, Alan Lowe

Main reference Christopher J. Soelistyo, Alan R. Lowe: “Discovering interpretable models of scientific image data with deep learning”, CoRR, Vol. abs/2402.03115, 2024.

URL <https://doi.org/10.48550/ARXIV.2402.03115>

Deep learning (DL) is now a powerful tool in microscopy data analysis, routinely used for image processing applications such as segmentation and denoising. However, it is rarely used to directly learn scientific models of a biological system, owing to the complexity of the internal representations. Here, we present our recent attempts to learn interpretable DL-based models of complex cell biological phenomena directly from a large corpus of time-lapse imaging data. In particular, we implement disentangled representation learning, causal time series models, network sparsity and symbolic methods, and assess their usefulness in forming interpretable models of complex data. We find that such methods can produce highly parsimonious models that achieve $\sim 98\%$ of the accuracy of black-box benchmark models, with a tiny fraction of the complexity. We explore the utility of such interpretable models in producing scientific explanations of the underlying biological phenomenon.

References

- 1 Soelistyo, Christopher and Vallardi, Giulia and Charras, Guillaume and Lowe, Alan. (2022) *Learning biophysical determinants of cell fate with deep neural networks*. Nature Machine Intelligence
- 2 Soelistyo, Christopher and Charras, Guillaume and Lowe, Alan. (2023) *Virtual perturbations to assess explainability of deep-learning based cell fate predictors*. In Proceedings of the IEEE/CVF International Conference on Computer Vision
- 3 Soelistyo, Christopher and Lowe, Alan. (2024) *Discovering interpretable models of scientific image data with deep learning*. arXiv preprint arXiv:2402.03115

3.8 Improving trustworthiness of ML in bioimaging through experimentally testable explanations

Axel Mosig (*Ruhr-Universität Bochum, DE*)

License © Creative Commons BY 4.0 International license
© Axel Mosig

Main reference David Schuhmacher, Stephanie Schörner, Claus Küpper, Frederik Großerüschkamp, Carlo Sternemann, Celine Lugnier, Anna-Lena Kraeft, Hendrik Jütte, Andrea Tannapfel, Anke Reinacher-Schick, Klaus Gerwert, Axel Mosig: “A framework for falsifiable explanations of machine learning models with an application in computational pathology”, *Medical Image Anal.*, Vol. 82, p. 102594, 2022.

URL <https://doi.org/10.1016/J.MEDIA.2022.102594>

The black box nature of neural networks is commonly regarded as the main source why predictions obtained from deep neural networks, despite their often unprecedented predictive accuracy, are often considered untrustworthy. In this contribution, I argue that the lack of trustworthiness of machine learning in general is due to its inductive nature: Machine learning models are obtained from inductive inferences, where specific observations in the form of training data are used to infer a general model that can classify data points beyond

the training data. From this perspective, machine learning is subject to the problem of induction, which has been brought to the point by the no-free-lunch theorem: Since there is no justification to assume that future events will resemble the past, all machine learning algorithms perform equal in terms of their out-of-training error.

Our further reasoning follows two interpretations, a global and a local interpretation, of the no-free-lunch theorem, which have been formulated recently by Sterkenburg and Grünwald. The global interpretation is in a sense the pessimistic interpretation, stating that no universal learning algorithm exists, since across the domain of all possible learning problems, all classifiers are identical in terms of their out-of-training error. The local interpretation is more constructive towards applied machine learning: When dealing with one specific problem, some learning algorithms do perform better on this specific task than other learning algorithms. This can be understood in terms of the inductive bias of different learning algorithms: As a direct implication of the no-free-lunch theorem, each learning algorithms must involve an either implicit or an explicit set of assumptions about how to generalize to data points beyond the training data. This set of assumptions is referred to as the inductive bias of a learning algorithm. From the perspective of one specific learning task, one can now ask what learning algorithm has an inductive bias that matches the underlying learning problem. This local interpretation of the no-free-lunch theorem essentially leads to considering machine learning as an inductive bias modeling problem.

The question that follows the local interpretation of the no-free-lunch theorem is how to justify inductive bias. I argue that our recently proposed framework for falsifiable explanations of artificial intelligence, or FXAI framework for short, addresses this question: The FXAI framework builds on the concept of explainability methods for neural networks, which usually provide an explainable output along with the classification of an input item. In the case of image classification, for example, the interpretable output is often a heat map that indicates which input variables have been relevant for obtaining the classification result of a specific image. In the FXAI framework, this explainable extension of the output is referred to as the interpretable space, or I-space for short. It is important to realize that an I-space, while being interpretable, can usually not be considered an interpretation in itself. The role of an interpretation (or, synonymously, an explanation) is rather assigned to a hypothesis that, in the sense of a scientific hypothesis, is required to be experimentally testable. The latter criterion is of crucial importance: Now, the explanation – and along with it, the I-space and hence the machine learning model – can be tested experimentally.

Experimental testability has relevant consequences: First of all, since the experiment that tests the explaining hypothesis is a different experiment than the experiment that yielded the data that were input to the machine learning model, the FXAI framework yields an experimental, deductive path to validate a machine learning model that is fully independent of cross validation. Second, the testable hypothesis suggests what should guide the inductive bias of a learning algorithm: namely an experimentally testable hypothesis.

We can now finally argue why experimentally testable explanations improve the trustworthiness of machine learning models. My argument lies in the nature of scientific hypotheses, which usually do not refer to one specific experiment. Rather, a strong hypothesis will usually suggest a wide range of different experiments through which the hypothesis can be tested. The more experiments a hypothesis invites for it to be tested, the more vulnerable the hypothesis becomes, because each experiment potentially falsifies the hypothesis. If, on the other hand, the hypothesis withstands all experimental attempts to falsify it, then the trustworthiness of the hypothesis and with it the associated machine learning model is undermined.

References

- 1 David Schuhmacher, Stephanie Schörner, Claus Küpper, Frederik Großerueschkamp, Carlo Sternemann, Celine Lugnier, Anna-Lena Kraeft, Hendrik Jütte, Andrea Tannapfel, Anke Reinacher-Schick, et al. A framework for falsifiable explanations of machine learning models with an application in computational pathology. *Medical Image Analysis*, 82:102594, 2022.
- 2 Tom F Sterkenburg and Peter D Grünwald. The no-free-lunch theorems of supervised learning. *Synthese*, 199(3-4):9979–10015, 2021.

3.9 Segment Anything for Microscopy

Constantin Pape (Universität Göttingen, DE)

License © Creative Commons BY 4.0 International license
© Constantin Pape

Main reference Anwai Archit, Sushmita Nair, Nabeel Khalid, Paul Hilt, Vikas Rajashekar, Marei Freitag, Sagnik Gupta, Andreas Dengel, Sheraz Ahmed, Constantin Pape: “Segment Anything for Microscopy”, bioRxiv, Cold Spring Harbor Laboratory, 2023.

URL <https://doi.org/10.1101/2023.08.21.554208>

The segmentation of cells in light microscopy or organelles in electron microscopy is one of the fundamental tasks in microscopy image analysis. While deep learning based approaches have improved segmentation qualities for a wide array of tasks, these solutions require specialized architectures and, unless very similar training data is publicly available, a significant amount of manual annotation. Recently versatile models that can be applied to a wider set of vision tasks – commonly referred to as foundation models – have been introduced. These models promise to bridge this gap and enable readily available solutions for many vision tasks. The foundation model “Segment Anything” developed by Meta implements this paradigm for segmentation tasks and can be applied for interactive and automatic segmentation in a large variety of image modalities. Our work builds on Segment Anything and evaluates and improves it for microscopy data. In particular, we implement a fine-tuning methodology that significantly improves the quality for microscopy and a software plugin for fast interactive data annotation., showing the promise of vision foundation models for microscopy image analysis.

3.10 High-volume, label-free imaging for quantifying single-cell dynamics in iPSC colonies

Anne Plant (NIST – Gaithersburg, US)

License © Creative Commons BY 4.0 International license
© Anne Plant

Joint work of Anthony Asmar, Zackery Benson, Adele Peskin, Mylene Simon, Michael Halter
Main reference Anthony Asmar, Zack Benson, Adele P. Peskin, Joe Chalfoun, Mylene Simon, Michael Halter, Anne Plant: “High-volume, label-free imaging for quantifying single-cell dynamics in induced pluripotent stem cell colonies”, bioRxiv, Cold Spring Harbor Laboratory, 2023.


URL <https://doi.org/10.1101/2023.09.29.558451>

To facilitate the characterization of unlabeled induced pluripotent stem cells (iPSCs) during culture and expansion, and to be able to address gene expression in individual living cells over time, we developed an AI pipeline for nuclear segmentation and mitosis detection from phase contrast images of individual cells within iPSC colonies. The analysis uses a 2D convolutional neural network (U-Net) plus a 3D U-Net applied on time lapse images to detect and segment

nuclei, mitotic events, and daughter nuclei to enable tracking of hundreds of thousands of individual cells over long times in culture. The analysis uses fluorescence data to train models for segmenting nuclei in phase contrast images. The use of classical image processing routines to segment fluorescent nuclei precludes the need for manual annotation and provides hundreds of thousands of cell objects for training. We explored reproducibility and generalizability of the pipeline, and how pipeline parameters influenced metrics of accuracy. The model is generalizable in that it performs well on different datasets with an average F1 score of 0.94, on cells at different densities, and on cells from different pluripotent cell lines. The method allows us to assess, in a non-invasive manner, rates of mitosis and cell division which serve as indicators of cell state and cell health. We assess these parameters in culture for more than 36 hours, at different locations in the colonies, and as a function of excitation light exposure.

3.11 Publishing microscopy and AI in Nature Methods

Rita Strack (Nature Publishing Group, US)

License  Creative Commons BY 4.0 International license
© Rita Strack

Reporting microscopy data and metadata are critical for data reproducibility, sharing, and reuse, and journals can have a key role in improving reporting standards. This talk discussed a methodological reporting crisis in microscopy, published works seeking to address this issue, and standards that are being implemented at Nature Methods. It also discussed the unique challenges associated with publishing reproducible AI for use in bioimage analysis and why this is crucial for the field moving forward. The goal was to inspire researchers to develop and implement best practice to promote reproducibility and growth within the field.

3.12 Frequency shortcuts learning and generalization in computer vision

Nicola Strisciuglio (University of Twente – Enschede, NL)

License  Creative Commons BY 4.0 International license
© Nicola Strisciuglio

Main reference Shunxin Wang, Raymond N. J. Veldhuis, Christoph Brune, Nicola Strisciuglio: “What do neural networks learn in image classification? A frequency shortcut perspective”, in Proc. of the IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023, pp. 1433–1442, IEEE, 2023.

URL <https://doi.org/10.1109/ICCV51070.2023.00138>

Neural networks trained through optimization techniques based on variants of stochastic gradient descent (SGD) present a spectral bias. Model training dynamics are biased towards learning features related to low-frequency components of the input data at early stages of training, and subsequently focusing on high-frequency features. Another important phenomenon in training dynamics is the emergence of shortcut learning, that is learning spurious correlations between the input data and prediction target. This results from the tendency of SGD-based training to find solutions that simplify the minimization of a loss function used as target of the training optimization problem. Shortcuts harm the generalization abilities of neural networks, especially in out-of-distribution (OOD) settings, and thus require particular attention during model validation.

We investigate the relationship between spectral bias and shortcut learning in image classification and expose the existence of shortcuts learned by vision models in the Fourier domain, which we call frequency shortcuts. We propose a method to detect possible frequency shortcuts, based on the importance that single frequency components have in the classification task, and construct dominant frequency maps (DFM). We demonstrate that frequency shortcuts can be learned at low or high-frequency and potentially harm the generalization capabilities in out-of-distribution settings, showing that shortcuts presents in OOD data can cause an illusion of strong generalization. In order to mitigate their impact on model performance, we also investigate the use of DFMs in a negative data augmentation strategy that improves adversarial robustness. However, extensive analysis of shortcuts learned by vision models is necessary and requires substantial attention to validate model performance and transferability to real-world tasks.

3.13 Visual interpretability of deep learning models in cell imaging

Assaf Zaritsky (Ben Gurion University – Beer Sheva, IL)

License © Creative Commons BY 4.0 International license
© Assaf Zaritsky

Joint work of Oded Rotem, Tamar Schwartz, Ron Maor, Yishay Tauber, Maya Tsarfati Shapiro, Marcos Meseguer, Daniella Gilboa, Daniel S. Seidman, Assaf Zaritsky

Main reference Oded Rotem, Tamar Schwartz, Ron Maor, Yishay Tauber, Maya Tsarfati Shapiro, Marcos Meseguer, Daniella Gilboa, Daniel S. Seidman, Assaf Zaritsky: “Visual interpretability of image-based classification models by generative latent space disentanglement applied to in vitro fertilization”, bioRxiv, Cold Spring Harbor Laboratory, 2023.

URL <https://doi.org/10.1101/2023.11.15.566968>

With the rapid growing volume and complexity of modern biomedical visual data, we can no longer rely on humans’ amazing capacity to identify visual patterns in biomedical images. Deep learning has emerged as a powerful technique to identify hidden patterns that exceed human intuition in complex cell imaging data. Extracting a deeper biological understanding, such as mechanistic description of complex phenotypes, require human interpretable explanation of the deep learning model’s decision process, however, the non-linear entanglement of image features makes deep learning models a “black box” that lacks straightforward explanations of which biologically meaningful image properties are important for the models’ decision. In my talk I presented a new generalized method toward systematic visual interpretability of deep learning image-based classification models that relies on counterfactual visual explanations using a disentangled latent representation. This method enables visually intuitive traversal of the latent space and we applied it to decipher blastocysts morphological quality properties in the context of in vitro fertilization.

3.14 Foundation models for biomedical image analysis

Shanghang Zhang (Peking University, CN)

License © Creative Commons BY 4.0 International license
© Shanghang Zhang

In this presentation, we delve into the potential of Foundation Models (FMs), which encompass Large Language Models (LLMs), Large Vision Models (VLMs), and Multimodal Large Language Models (MLLM). These models have demonstrated promising outcomes in various

scenarios. However, integrating FM capabilities into professional domains remains an unresolved inquiry. We present some recent relevant research endeavours to address emergent challenges during this transition. The initial query pertains to efficiently aligning data from specialized domains with non-specific FMs. Parameter Efficient Fine-tuning (PEFT) offers a viable approach, and to adapt PEFT more suitably for medical data in our case, we have devised a tree-like structured adapter that hierarchically incorporates medical knowledge into the Segment Anything (SAM) model. Secondly, we illustrate how quantization techniques can accelerate FM performance for biological tasks. Subsequently, we demonstrate the fine-tuning process of an MLLM using medical data to generate medical reports and accomplish vision-based question-answering tasks. This process leverages methods such as in-context learning to align training data across different modalities with retrieval augmentative generation to support our model giving a more comprehensive report.

3.15 Algorithmic Fairness, Robust Generalization and Trustworthy Machine Learning

Han Zhao (University of Illinois – Urbana-Champaign, US)

License  Creative Commons BY 4.0 International license
© Han Zhao

Joint work of Han Zhao, Haoxiang Wang, Haozhe Si, Gargi Balasubramaniam, Bo Li

In this talk, I will discuss two important aspects of machine learning: algorithmic fairness and robust generalization under the common framework of invariant causal prediction. I will first provide some motivating examples of these two problems in the context of biomedical and healthcare applications. I will then introduce our recent work [1, 2] on invariant feature recovery to address the above two problems. I will conclude the talk with a discussion of some open problems and future research directions.

References

- 1 Wang, Haoxiang and Si, Haozhe and Li, Bo and Zhao, Han. *Provable domain generalization via invariant-feature subspace recovery*. In Proceedings of the 39th International Conference on Machine Learning (ICML 2022)
- 2 Wang, Haoxiang and Balasubramaniam, Gargi and Si, Haozhe and Li, Bo and Zhao, Han. *Invariant-Feature Subspace Recovery: A New Class of Provable Domain Generalization Algorithms*. arXiv preprint arXiv:2311.00966

Participants

- Chao Chen
Stony Brook University, US
- Jianxu Chen
ISAS – Dortmund, DE
- Evangelia Christodoulou
DKFZ – Heidelberg, DE
- Beth Cimini
Broad Institute of MIT & Harvard – Cambridge, US
- Gaole Dai
Peking University, CN
- Meghan Driscoll
University of Minnesota – Minneapolis, US
- Edward Evans III
University of Wisconsin – Madison, US
- Matthias Gunzer
Universität Duisburg-Essen, DE & ISAS e.V. – Dortmund, DE
- Andrew Hufton
Patterns, Cell Press – Würzburg, DE
- Florian Jug
Human Technopole – Milano, IT
- Anna Kreshuk
EMBL – Heidelberg, DE
- Thomas Lemberger
EMBO – Heidelberg, DE
- Alan Lowe
The Alan Turing Institute – London, GB
- Shalin Mehta
Chan Zuckerberg Biohub – Stanford, US
- Axel Mosig
Ruhr-Universität Bochum, DE
- Matheus Palhares Viana
Allen Institute for Cell Science – Seattle, US
- Constantin Pape
Universität Göttingen, DE
- Anne Plant
NIST – Gaithersburg, US
- Susanne Rafelski
Allen Institute for Cell Science – Seattle, US
- Ananya Rastogi
Springer Nature – New York, US
- Albert Sickmann
ISAS – Dortmund, DE
- Rita Strack
Nature Publishing Group, US
- Nicola Strisciuglio
University of Twente – Enschede, NL
- Aubrey Weigel
Howard Hughes Medical Institute – Ashburn, US
- Assaf Zaritsky
Ben Gurion University – Beer Sheva, IL
- Shanghang Zhang
Peking University, CN
- Han Zhao
University of Illinois – Urbana-Champaign, US



Next Generation Protocols for Heterogeneous Systems

Stephanie Balzer^{*1}, Marco Carbone^{*2}, Roland Kuhn^{*3}, and Peter Thiemann^{*4}

1 Carnegie Mellon University, USA. balzers@cs.cmu.edu

2 IT University of Copenhagen, DK. carbonem@itu.dk

3 Actyx AG – München, DE. roland@actyx.io

4 University of Freiburg, DE. thiemann@informatik.uni-freiburg.de

Abstract

The emergence of new computing systems, like cloud computing, blockchains, and Internet of Things (IoT), replaces the traditional monolithic software hardware stack with a distributed heterogeneous model. This change poses new demands on the programming languages for developing such systems: *compositionality*, allowing decomposition of a system into smaller, possibly heterogeneous, parts and composition of the individually verified parts into a verified whole, *security*, asserting end-to-end integrity and confidentiality, *quantitative reasoning methods*, accounting for timing and probabilistic events, and, as a cross-cutting concern, *certification* of asserted properties in terms of independently verifiable, machine-checked proofs.

Characteristics of this emerging computation model are distribution of the participating entities and message passing as the primary means of communication. Message passing is also the communication model underlying behavioral types and programming languages, making them uniquely fitted for this new application domain. Behavioral types explicitly capture the protocols of message exchange and have a strong theoretical foundation. Recent applications of behavioral types include smart contract languages, information flow control, and machine-checked proofs of safety properties. Although these early explorations are promising, the current state of the art of behavioral types and programming languages lacks a comprehensive account of the above-mentioned demands.

This Dagstuhl Seminar aims to gather experts from academia and industry to discuss the use of programming languages tailored to tackle the challenges posed by today's emerging distributed and heterogeneous computing platforms, e.g., by making use of behavioral types. It will focus on static and possibly dynamic mechanisms to support compositionality, security, quantitative reasoning, and certification.

Seminar January 28 – February 2, 2024 – <https://www.dagstuhl.de/24051>

2012 ACM Subject Classification Theory of computation → Process calculi; Theory of computation → Type structures

Keywords and phrases behavioural types, concurrency, programming languages, session types

Digital Object Identifier 10.4230/DagRep.14.1.108

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Next Generation Protocols for Heterogeneous Systems, *Dagstuhl Reports*, Vol. 14, Issue 1, pp. 108–129

Editors: Stephanie Balzer, Marco Carbone, Roland Kuhn and Peter Thiemann



Dagstuhl Reports

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

1 Executive Summary

Stephanie Balzer (Carnegie Mellon University, USA)

Marco Carbone (IT University of Copenhagen, DK)

Roland Kuhn (Actyx AG – München, DE)

Peter Thiemann (University of Friburg, DE)

License © Creative Commons BY 4.0 International license
© Stephanie Balzer, Marco Carbone, Roland Kuhn, and Peter Thiemann

This Dagstuhl Seminar followed the earlier Dagstuhl Seminars 17051 “Theory and Applications of Behavioural Types” and 21372 “Behavioural Types: Bridging Theory and Practice”. Whereas Seminar 17051 was focusing on theoretical aspects of behavioural types, and Seminar 21372 focused on bridging the gap with practical application, this seminar was much broader and aimed at extending to other communities such as security and other areas of programming languages.

Initial preparations

Based on the ideas of our seminar proposal, we established four key general areas: quantitative systems, verification, mechanisation, and security. We assigned each area to a day of the week (from Monday to Thursday) and asked an invitee representative of the area to give an introductory talk. Then, each of these talks was followed by other talks and breakout rooms related to the area. Breakout rooms were established during the seminar based on discussions with the rest of the participants. As a result of this, the first part of the week consisted primarily of talks, while the second part included more time for breakout sessions.

Activities and outcomes

Throughout the seminar, the participants gathered in focused breakout groups: the findings of the breakout groups are described in more detail in the last part of the report. The participants of several breakout groups have agreed to continue their work and collaboration after the seminar.

In addition to these more structured breakout sessions there were further lively improvised meetings and discussions (especially after dinner) which are not summarised in the report.

Overall, we believe that the seminar activities were a success. At the end of the seminar the participants agreed to remain in contact to continue the discussions, and foster new collaborations. There was strong enthusiasm for organising a follow-up Dagstuhl Seminar in the future, perhaps taking place in about 1–2 years time. One concrete outcome was the submission of a position paper (cf. the working group “Typing Across Heterogeneous Components”) that has been accepted and presented at PLACES 2024 (co-located with ETAPS).

2 Table of Contents

Executive Summary

Stephanie Balzer, Marco Carbone, Roland Kuhn, and Peter Thiemann 109

Overview of Talks


Area talk: Security of Heterogenous Systems: Principles, Practice, and a Case for Secure Runtimes <i>Aslan Askarov</i>	112
Logical Relations for Session-Typed Concurrency <i>Stephanie Balzer</i>	112
Area talk: Program Development Tools for Secure Multi-Party Computation <i>Marina Blanton</i>	112
Contracts for Session-based Programming with Linear Dependent Types <i>Luis Caires</i>	113
Regrading Policies for Flexible Information Flow Control in Session-Typed Concurrency <i>Farzaneh Derakhshan</i>	113
The Rational Programmer <i>Christos Dimoulas</i>	114
Special Delivery: Programming with Mailbox Types <i>Simon Fowler</i>	114
Correct orchestration of Federated Learning: formalisation and verification <i>Silvia Ghilezan</i>	115
Information-Flow Control in Choreographies <i>Andrew Hirsch</i>	115
Actris tool presentation <i>Jonas Kastberg Hinrichsen</i>	115
Area talk: Mechanized verification of type systems using Iris <i>Robbert Krebbers</i>	116
Behavioural Types for Local-First Software: replicated roles, full availability <i>Roland Kuhn</i>	116
Probabilistic Theories of Choreographic Programming <i>Marco Peressotti</i>	117
Comparing Process Calculi using Encodings <i>Kirstin Peters</i>	117
Mechanizing Session-Types: Enforcing linearity without linearity <i>Brigitte Pientka</i>	118
Router-based Analysis of Multiparty Protocols <i>Jorge Pérez</i>	118
Behavioural up/down casting for statically typed languages (tool presentation) <i>António Ravara</i>	119

Deciding Subtyping for Asynchronous Multiparty Sessions <i>Felix Stutz</i>	119
Area talk: Quantitative techniques <i>Emilio Tuosto</i>	120
System f_{ω}^{μ} with context-free session types <i>Vasco T. Vasconcelos</i>	120
STL3: Toward Security via Free Theorems in a Session-Typed Linear Language with Locations <i>Andrew Wagner</i>	120
Area talk: Verification <i>Nobuko Yoshida</i>	121
Working groups	
Breakout Group: IFC and Noninterference <i>Aslan Askarov, Stephanie Balzer, Marina Blanton, Christos Dimoulas, Emanuele D’Osualdo, Farzaneh Derakhshan, Andrew Wagne</i>	121
Breakout Group: Secure Multiparty Computation <i>Amal Ahmed, Aslan Askarov, Stephanie Balzer, Andrew Wagner, Marina Blanton, Christos Dimoulas, Emanuele D’Osualdo, Farzaneh Derakhshan, Philipp Haller</i>	121
Breakout Group: Logical Relations and Session Types <i>Amal Ahmed, Stephanie Balzer, Luis Caires, Emanuele D’Osualdo, Farzaneh De- rakhshan, Adrian Francalanza, Ralf Jung, Robbert Krebbers, Peter Thiemann, Andrew Wagner</i>	122
Real-World Applications of Behavioural Types <i>Kirstin Peters, Silvia Ghilezan, Jesper Bengtson, Christos Dimoulas, Marco Car- bone, Felix Stutz, Antonio Ravara</i>	122
Typing Across Heterogeneous Components <i>Roland Kuhn, Philipp Haller, Sam Lindley, Vasco T. Vasconcelos, Simon Fowler, Alceste Scalas, Malte Viering, Raymond Hu</i>	123
Probabilistic Behavioural Types <i>Emilio Tuosto, Silvia Ghilezan, Emanuele D’Osualdo, Jorge Pérez, Nobuko Yoshida, Marco Carbone, Marco Peressotti, Kirstin Peters, Alceste Scalas</i>	124
Open World Choreographies <i>Andrew Hirsch, Lukasz Ziarek, Marco Peressotti, Malte Viering, Raymond Hu, Roland Kuhn</i>	125
Mechanisation of Behavioural Types <i>Jesper, Luis, Kirstin, Robbert, Jonas, Alceste, Ralf</i>	126
Dependent Session Types	127
Participants	129

3 Overview of Talks

3.1 Area talk: Security of Heterogenous Systems: Principles, Practice, and a Case for Secure Runtimes


Aslan Askarov (Aarhus University – Aarhus, Denmark)

License  Creative Commons BY 4.0 International license
 © Aslan Askarov

The classical computer security principle of the least common mechanism says that resource sharing creates security problems and should be treated carefully. This talk highlights that programming language runtimes that handle sensitive information at different confidentiality levels are such common mechanisms and, therefore, can inadvertently leak information. We examine runtime aspects such as schedulers and mailboxes and also study mitigating traffic analysis attacks using information flow techniques.

3.2 Logical Relations for Session-Typed Concurrency


Stephanie Balzer (Carnegie Mellon University – Pittsburgh, US)

License  Creative Commons BY 4.0 International license
 © Stephanie Balzer
 Joint work of Stephanie Balzer, Farzaneh Derakhshan, Robert Harper, Yue Yao

Program *equivalence* is the fulcrum for reasoning about and proving properties of programs. For noninterference, for example, program equivalence up to the secrecy level of an observer is shown. A powerful enabler for such proofs are *logical relations*. Logical relations only were adopted for session types relatively recently – but exclusively for terminating languages. This talk scales logical relations to *general recursive session types*. It develops a logical relation for progress-sensitive noninterference (PSNI) for intuitionistic linear logic session types (ILLST), tackling the challenges non-termination and concurrency pose, and shows that logical equivalence is *sound and complete* with regard to closure of weak bisimilarity under parallel composition, using a *biorthogonality* argument. A distinguishing feature of the logical relation is its stratification with an *observation index* (as opposed to a step or unfolding index), a crucial shift to make the logical relation closed under parallel composition in a concurrent setting.

3.3 Area talk: Program Development Tools for Secure Multi-Party Computation

Marina Blanton (University at Buffalo – Buffalo, US)


License  Creative Commons BY 4.0 International license
 © Marina Blanton

Secure multi-party computation permits evaluation of a function or a program on protected private inputs in such a way that the computation participants have no access to the data in the clear throughout the computation. The security guarantees are such that only the computation outcome becomes disclosed to the designated parties. In this talk, we discussed the setup, security definitions, and their differences from the properties of other definitions

used in the programming languages community. The second part of the talk discussed PICCO, a compiler for transforming general-purpose programs intended to be executed on private data into the corresponding secure multi-party computation protocols. We discussed compiler optimizations and mechanisms for making it easier for programmers to develop efficient programs in this framework.

3.4 Contracts for Session-based Programming with Linear Dependent Types


Luis Caires (IST – Lisbon, PT)

License  Creative Commons BY 4.0 International license
© Luis Caires

We sketch a novel approach to linear dependent session types based on a Proposition-as-Types foundation of session-based programs, which targets the development of a linear dependent type theory with equality types for session behaviour, allowing properties of linear objects to be expressed.

3.5 Regrading Policies for Flexible Information Flow Control in Session-Typed Concurrency

Farzaneh Derakhshan (Illinois Institute of Technology – Chicago, US)

License  Creative Commons BY 4.0 International license
© Farzaneh Derakhshan
Joint work of Farzaneh Derakhshan, Stephanie Balzer, Yue Yao

Noninterference guarantees that an attacker cannot infer secrets by interacting with a program. An information flow control type system asserts noninterference by tracking the level of information learned and disallowing leakage to entities of lesser or unrelated levels. These restrictions cater to scenarios in which the information learned by an entity monotonically increases with program progression but are at odds with control flow constructs, permitting interaction with entities of lower levels in the continuation. Relaxing such restrictions is particularly challenging in a concurrent setting. This paper utilizes session types to track the flow of information and develops an information flow control type system for message-passing concurrent processes that allows downgrading the pc for the next loop iteration upon recursion. To ensure noninterference, the type system relies on regrading policies, ensuring that any confidential information learned during the high-security parts of the loop cannot be rolled forward to the next iteration. To express regrading policies, the type system is complemented with integrity to ensure that entities with different regrading policies can be safely composed. The paper develops the type system and proves progress-sensitive noninterference for well-typed programs, ruling out timing attacks that exploit the relative order of messages. The type system has been implemented in a type checker, which supports security-polymorphic processes using local security theories.

3.6 The Rational Programmer

Christos Dimoulas (Northwestern University – Evanston, US)

License  Creative Commons BY 4.0 International license
© Christos Dimoulas

The productivity of developers depends on the quality of the available programming languages: their support for adequate testing, for locating and fixing mistakes, and for maintenance tasks. If a programming language does not support these routine tasks, the developer is forced to resort to labor-intensive and ineffective workarounds. Put differently, it isn't about the syntax, the types, or the semantics, but about the pragmatics of a programming language. The problem is that the PL research area has so far few tools to evaluate pragmatics.

The Rational Programmer is a new scientific instrument for that purpose. While simulations have a long history in computer science applications, the Rational Programmer method puts them to new use in PL research. The heart of the method is a simulation, namely an algorithmic abstraction of information gathering in a work context. The outcome of a rational programmer simulation is typically a strategy that a developer can employ. It may also point designers and researchers to a problematic aspect of a language. Finally, it can inform instructors how to teach students the effective use of a language. In this talk, I will demonstrate the workings of the Rational Programmer method with examples.

3.7 Special Delivery: Programming with Mailbox Types

Simon Fowler (University of Glasgow, GB)

License  Creative Commons BY 4.0 International license
© Simon Fowler

The asynchronous and unidirectional communication model supported by mailboxes is a key reason for the success of actor languages like Erlang and Elixir for implementing reliable and scalable distributed systems. While many actors may send messages to some actor, only the actor may (selectively) receive from its mailbox. Although actors eliminate many of the issues stemming from shared memory concurrency, they remain vulnerable to communication errors such as protocol violations and deadlocks.

Mailbox types are a novel behavioural type system for mailboxes first introduced for a process calculus by de'Liguoro and Padovani in 2018, which capture the contents of a mailbox as a commutative regular expression. Due to aliasing and nested evaluation contexts, moving from a process calculus to a programming language is challenging. This paper presents Pat, the first programming language design incorporating mailbox types, and describes an algorithmic type system. We make essential use of quasi-linear typing to tame some of the complexity introduced by aliasing. Our algorithmic type system is necessarily co-contextual, achieved through a novel use of backwards bidirectional typing, and we prove it sound and complete with respect to our declarative type system. We implement a prototype type checker, and use it to demonstrate the expressiveness of Pat on a factory automation case study and a series of examples from the Savina actor benchmark suite.

3.8 Correct orchestration of Federated Learning: formalisation and verification

Silvia Ghilezan (Mathematical Institute – Belgrade, RS)

License © Creative Commons BY 4.0 International license
© Silvia Ghilezan

Main reference Ivan Prokic, Silvia Ghilezan, Simona Kasterovic, Miroslav Popovic, Marko Popovic, Ivan Kastelan: “Correct orchestration of Federated Learning generic algorithms: formalisation and verification in CSP”, CoRR, Vol. abs/2306.14529, 2023.

URL <https://doi.org/10.48550/ARXIV.2306.14529>

Federated learning (FL) is a machine learning setting where clients keep the training data decentralised and collaboratively train a model either under the coordination of a central server (centralised FL) or in a peer-to-peer network (decentralised FL). Correct orchestration is one of the main challenges. In this paper, we formally verify the correctness of two generic FL algorithms, a centralised and a decentralised one, using the CSP process calculus and the PAT model checker. The CSP models consist of CSP processes corresponding to generic FL algorithm instances. PAT automatically proves the correctness of the two generic FL algorithms by proving their deadlock freeness (safety property) and successful termination (liveness property). The CSP models are constructed bottom-up by hand as a faithful representation of the real Python code and is automatically checked top-down by PAT.

3.9 Information-Flow Control in Choreographies

Andrew Hirsch (University at Buffalo – SUNY, US)

License © Creative Commons BY 4.0 International license
© Andrew Hirsch

Information-flow control is an important information-security-enforcement mechanism. It requires that secret information not be allowed to influence (or be used to compute) public data. This ensures that no private data will be leaked to the outside world. However, enforcing information-flow control in the concurrent world has proven incredibly difficult. The only known versions are incredibly restrictive. In current work, we are adding information-flow control to choreographic programs, where the extra restrictiveness is not necessary. In this talk, I will explain what goes wrong with information-flow control in concurrent settings, and why choreographic programming appears to rescue it.

3.10 Actris tool presentation

Jonas Kastberg Hinrichsen (Aarhus University, DK)

License © Creative Commons BY 4.0 International license
© Jonas Kastberg Hinrichsen
URL <https://iris-project.org/actris/>

Binary sessions, specifying bidirectional exchanges of messages between two processes, has widely been used to model idealised reliable communication, where messages are never dropped, duplicated, or arrive out of order. Such sessions allow for sophisticated protocol structures, as evidenced by the ever expanding work on session types, a behavioural type

system for specifying the types of individual messages of a sequence of exchanges. However, the expressivity of session types is often restricted to a decidable fragment, which prohibits them from reasoning about functional correctness.

In this tool presentation, I present the ongoing story of Actris tool – a framework for session type-based reasoning in separation logic – and how it can be used to reason about reliable communication. In particular, I demonstrate the full verification of a suite of programs that combine message passing and shared memory concurrency. I additionally briefly cover the various extensions of Actris; notably how it has been applied to the verification of distributed systems and deadlock freedom.

3.11 Area talk: Mechanized verification of type systems using Iris

Robbert Krebbers (Radboud University Nijmegen, NL)

License  Creative Commons BY 4.0 International license
© Robbert Krebbers

This talk gives an introduction to the “logical approach” to proving type safety. I will first present a simple version, and then scale up to a small session-typed language. I will show that this approach is well-suited for mechanization of challenging type systems in the Coq proof assistant

3.12 Behavioural Types for Local-First Software: replicated roles, full availability

Roland Kuhn (Actyx AG)

License  Creative Commons BY 4.0 International license
© Roland Kuhn

Joint work of Roland Kuhn, Hernán Melgratte, Emilio Tuosto

Main reference Roland Kuhn, Hernán C. Melgratti, Emilio Tuosto: “Behavioural Types for Local-First Software”, in Proc. of the 37th European Conference on Object-Oriented Programming, ECOOP 2023, July 17-21, 2023, Seattle, Washington, United States, LIPIcs, Vol. 263, pp. 15:1–15:28, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023.

URL <https://doi.org/10.4230/LIPICSECOOP.2023.15>

In this work we formalise an existing system for high-availability industry automation: the constraint that availability must be maximised – even at the cost of strong consistency – poses some interesting challenges. The basis is given by the peer-to-peer, uncoordinated, but causality-preserving event log replication of the Actyx middleware. Our work resulted in well-formedness constraints that ensure that a designed interaction protocol will achieve eventual consensus on the event trace of its execution, without any need for further coordination. This holds even in a dynamic swarm setting, where any role can be replicated any number of times and new participants can join and leave the system at any time.

3.13 Probabilistic Theories of Choreographic Programming

Marco Peressotti (University of Southern Denmark – Odense, DK)

License © Creative Commons BY 4.0 International license
© Marco Peressotti

Choreographic programming is a paradigm for developing concurrent and distributed systems, where programs are choreographies that define, from a global viewpoint, the computations and interactions that communicating processes should enact. Choreography compilation translates choreographies into the local definitions of process behaviours, given as terms in a process calculus. In this talk we present the first theory of choreographic programming language that incorporates probabilistic aspects for local computation, choreographic choice, and scheduling. We start from an established theory of choreographic programming [1, 2, 3, 4, 5] and integrate various probabilistic features while maintaining the original syntax. We show that the original compilation procedure can still be used and establish its correctness in terms of probabilistic bisimilarity via standard up-to techniques. We discuss how the various probabilistic features impact the design of the semantics and the lessons learned while integrating them.

References

- 1 Montesi, F. 2023. Introduction to Choreographies. Cambridge University Press. DOI:10.1017/9781108981491
- 2 Cruz-Filipe Luís, Montesi, F. and Peressotti, M. 2023. A Formal Theory of Choreographic Programming. *Journal of Automated Reasoning*. 67, 21 (2023), 1–34. DOI:10.1007/s10817-023-09665-3.
- 3 Cruz-Filipe Luís, Montesi, F. and Peressotti, M. 2021. Certifying Choreography Compilation. *Theoretical Aspects of Computing – ICTAC 2021 – 18th International Colloquium, Virtual Event, Nur-Sultan, Kazakhstan, September 8-10, 2021, Proceedings (2021)*, 115–133. DOI:10.1007/978-3-031-17715-6_15
- 4 Cruz-Filipe Luís, Montesi, F. and Peressotti, M. 2021. Formalising a Turing-Complete Choreographic Language in Coq. *12th International Conference on Interactive Theorem Proving (ITP 2021) (Dagstuhl, Germany, 2021)*, 15:1–15:18. DOI:10.4230/LIPIcs.ITP.2021.15
- 5 Cruz-Filipe Luís, Graversen, E., Montesi, F. and Peressotti, M. 2023. Reasoning About Choreographic Programs. *Coordination Models and Languages (2023)*, 144–162. DOI:10.1007/978-3-031-35361-1_8

3.14 Comparing Process Calculi using Encodings

Kirstin Peters (Universität Augsburg, DE)

License © Creative Commons BY 4.0 International license
© Kirstin Peters

Encodings are often used to compare process calculi. To rule out trivial or meaningless encodings, they are augmented with encodability criteria. This talk is about how to reason about the quality of encodability criteria and how to set up such criteria.

3.15 Mechanizing Session-Types: Enforcing linearity without linearity

Brigitte Pientka (McGill University – Montréal, CA)

License  Creative Commons BY 4.0 International license
© Brigitte Pientka

Joint work of Brigitte Pientka, Chuta Sano, Ryan Kavanagh

Main reference Chuta Sano, Ryan Kavanagh, Brigitte Pientka: “Mechanizing Session-Types using a Structural View: Enforcing Linearity without Linearity”, CoRR, Vol. abs/2309.12466, 2023.

URL <https://doi.org/10.48550/ARXIV.2309.12466>

Session types employ a linear type system that ensures that communication channels cannot be implicitly copied or discarded. As a result, many mechanizations of these systems require modeling channel contexts and carefully ensuring that they treat channels linearly. We demonstrate a technique that localizes linearity conditions as additional predicates embedded within type judgments, which allows us to use structural typing contexts instead of linear ones. This technique is especially relevant when leveraging (weak) higher-order abstract syntax to handle channel mobility and the intricate binding structures that arise in session-typed systems.

Following this approach, we mechanize a session-typed system based on classical linear logic and its type preservation proof in the proof assistant Beluga, which uses the logical framework LF as its encoding language. We also prove adequacy for our encoding. This shows the tractability and effectiveness of our approach in modelling substructural systems such as session-typed languages.

3.16 Router-based Analysis of Multiparty Protocols

Jorge Pérez (University of Groningen, NL)

License  Creative Commons BY 4.0 International license
© Jorge Pérez

We are interested in the rigorous verification of message-passing programs, which operate by exchanging messages across distributed networks. Ensuring that these communicating programs are correct is important but highly challenging.

Originated from the realms of Concurrency Theory and Programming Languages, Multiparty Session Types (MPSTs) offer a convenient methodology for the development and verification of message-passing programs. The methodology of MPSTs offers a structured approach to the design of advanced verification techniques, both static (via type systems) and dynamic (via monitoring architectures). Interestingly, these static and verification techniques can be defined by following principled approaches based on resource-aware logics, in particular Girard’s Linear Logic.

In this talk, I will overview recent work by my group in this direction, and in particular I discuss how the concept of router of a multiparty protocol can be effective for runtime verification.

3.17 Behavioural up/down casting for statically typed languages (tool presentation)

António Ravara (NOVA University of Lisbon, PT)

License © Creative Commons BY 4.0 International license
 © António Ravara
URL <https://github.com/jdmota/java-typestate-checker>

We provide support for polymorphism in static typestate analysis for object-oriented languages with upcasts and downcasts. Recent work has shown how typestate analysis can be embedded in the development of Java programs to obtain safer behaviour at runtime, e.g., absence of null pointer errors and protocol completion. In that approach, inheritance is supported at the price of limiting casts in source code, thus only allowing those at the beginning of the protocol, i.e., immediately after objects creation, or at the end, and in turn seriously affecting the applicability of the analysis.

We provide a solution to this open problem in typestate analysis by introducing a theory based on a richer data structure, named typestate tree, which supports upcast and downcast operations at any point of the protocol by leveraging union and intersection types. The soundness of the typestate tree-based approach has been mechanised in Coq. The theory can be applied to most object-oriented languages statically analysable through typestates, thus opening new scenarios for acceptance of programs exploiting inheritance and casting. To defend this thesis, we show an application of the theory, by embedding the typestate tree mechanism in a Java-like object-oriented language, and proving its soundness.

Accepted in ECOOP'24.

3.18 Deciding Subtyping for Asynchronous Multiparty Sessions

Felix Stutz (University of Luxembourg, LU)

License © Creative Commons BY 4.0 International license
 © Felix Stutz
Joint work of Elaine Li, Felix Stutz, Thomas Wies
Main reference Elaine Li, Felix Stutz, Thomas Wies: “Deciding Subtyping for Asynchronous Multiparty Sessions”, in Proc. of the Programming Languages and Systems – 33rd European Symposium on Programming, ESOP 2024, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2024, Luxembourg City, Luxembourg, April 6-11, 2024, Proceedings, Part I, Lecture Notes in Computer Science, Vol. 14576, pp. 176–205, Springer, 2024.
URL https://doi.org/10.1007/978-3-031-57262-3_8

Multiparty session types (MSTs) are a type-based approach to verifying communication protocols, represented as global types in the framework. We present a precise subtyping relation for asynchronous MSTs with communicating state machines (CSMs) as implementation model. We address two problems: when can a local implementation safely substitute another, and when does an arbitrary CSM implement a global type? We define safety with respect to a given global type, in terms of subprotocol fidelity and deadlock freedom. Our implementation model subsumes existing work which considers local types with restricted choice. We exploit the connection between MST subtyping and refinement to formulate concise conditions that are directly checkable on the candidate implementations, and use them to show that both problems are decidable in polynomial time. This talk was given by Felix Stutz (Max Planck Institute for Software Systems). It is based on joint work with Elaine Li and Thomas Wies (New York University). In April 2024, the corresponding paper was published in the proceedings of the 33rd European Symposium on Programming (ESOP24).

3.19 Area talk: Quantitative techniques


Emilio Tuosto (Gran Sasso Science Institute – L’Aquila, IT)

License  Creative Commons BY 4.0 International license
© Emilio Tuosto

This talk gives a bird-eye watch of the literature at the intersection between quantitative techniques and behavioural specifications. More precisely, following the chronological order, the talk surveys the papers concerned with resource awareness, time, and probabilities. The talk strives to succinctly highlight the main contributions of each paper and distill some interesting open problems. Although related to the topic of the talk, data-awareness was intentionally left out of the exposition in order to maintain the focus centred on quantitative approaches.

3.20 System f_{ω}^{μ} with context-free session types

Vasco T. Vasconcelos (University of Lisbon, PT)

License  Creative Commons BY 4.0 International license
© Vasco T. Vasconcelos

Joint work of Diogo Poças, Diana Costa, Andreia Mordido, Vasco T. Vasconcelos

Main reference Diogo Poças, Diana Costa, Andreia Mordido, Vasco T. Vasconcelos: “System F_{ω}^{μ} with Context-free Session Types”, in Proc. of the Programming Languages and Systems – 32nd European Symposium on Programming, ESOP 2023, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2023, Paris, France, April 22-27, 2023, Proceedings, Lecture Notes in Computer Science, Vol. 13990, pp. 392–420, Springer, 2023.

URL https://doi.org/10.1007/978-3-031-30044-8_15

We study increasingly expressive type systems, from F^{μ} –an extension of the polymorphic lambda calculus with equirecursive types– to F_{ω}^{μ} –the higher-order polymorphic lambda calculus with equirecursive types and context-free session types. Type equivalence is given by a standard bisimulation defined over a novel labelled transition system for types. Our system subsumes the contractive fragment of F_{ω}^{μ} as studied in the literature. Decidability results for type equivalence of the various type languages are obtained from the translation of types into objects of an appropriate computational model: finite-state automata, simple grammars and deterministic pushdown automata. We show that type equivalence is decidable for a significant fragment of the type language. We further propose a message-passing, concurrent functional language equipped with the expressive type language and show that it enjoys preservation and absence of runtime errors for typable processes.

3.21 STL3: Toward Security via Free Theorems in a Session-Typed Linear Language with Locations

Andrew Wagner (Northeastern University – Boston, US)

License  Creative Commons BY 4.0 International license
© Andrew Wagner

Joint work of Andrew Wagner, Amal Ahmed

We present work in progress on a minimal extension to intuitionistic binary session types that reifies channel names into types. Along with quantification over names, this establishes a form of name parametricity with which common security properties like authenticity, binding,

and hiding can be expressed as free theorems. We identify some of the key challenges in proving these free theorems, and more specifically where standard techniques seem to be insufficient.

3.22 Area talk: Verification

Nobuko Yoshida (University of Oxford, GB)


License  Creative Commons BY 4.0 International license
© Nobuko Yoshida

I give the overview on verification of session types: (1) Linear Logic-Based Session Types; (2) Full Abstraction Results of System F; (3) Asynchronous Communication Optimisations in Rust; (4) Comparison of Performances in Session-Based Rust; and (5) Correspondence with Communication Automata.

4 Working groups

4.1 Breakout Group: IFC and Noninterference


Aslan Askarov, Stephanie Balzer, Marina Blanton, Christos Dimoulas, Emanuele D’Oswaldo, Farzaneh Derakhshan, and Andrew Wagner

License  Creative Commons BY 4.0 International license
© Aslan Askarov, Stephanie Balzer, Marina Blanton, Christos Dimoulas, Emanuele D’Oswaldo, Farzaneh Derakhshan, Andrew Wagner

During this break-out session, the discussion focused on the following points: state of the art in information flow control, quantitative information flow and its connection to secure multi-party computation, modern formulation of noninterference policies using the epistemic (knowledge-based) framework for declassification, and comparison of different noninterference theorem statements.

4.2 Breakout Group: Secure Multiparty Computation

Amal Ahmed, Aslan Askarov, Stephanie Balzer, Andrew Wagner, Marina Blanton, Christos Dimoulas, Emanuele D’Oswaldo, Farzaneh Derakhshan, and Philipp Haller

License  Creative Commons BY 4.0 International license
© Amal Ahmed, Aslan Askarov, Stephanie Balzer, Andrew Wagner, Marina Blanton, Christos Dimoulas, Emanuele D’Oswaldo, Farzaneh Derakhshan, Philipp Haller

During the session, we discussed the security definitions employed in the secure multi-party computation literature, properties of the resulting protocols, and their relationship to the definitions and properties achieved in other settings, e.g., in information flow control literature. Because the model provides powerful guarantees at the level of individual operations, it becomes possible to achieve properties which otherwise would be difficult to achieve in other settings. For example, the fact that a computation participant does not see private values implies that the participant is unable to leak secret data. During the discussion we were also seeking connections with other PL concepts and recent developments as a way to build a fruitful collaboration. Topics that came up include concurrency, information flow, and orchestration.

4.3 Breakout Group: Logical Relations and Session Types

Amal Ahmed, Stephanie Balzer, Luis Caires, Emanuele D’Oswaldo, Farzaneh Derakhshan, Adrian Francalanza, Ralf Jung, Robbert Krebbers, Peter Thiemann, and Andrew Wagner

License © Creative Commons BY 4.0 International license
 © Amal Ahmed, Stephanie Balzer, Luis Caires, Emanuele D’Oswaldo, Farzaneh Derakhshan, Adrian Francalanza, Ralf Jung, Robbert Krebbers, Peter Thiemann, Andrew Wagner

This breakout session gathered people currently employing the advanced proof method of logical relations and people interested in learning more about it. Of particular concern was the recent developments that employ logical relations in a session-typed concurrent setting. As such the session allowed experts to provide an overview of the current state of the art and latest achievements. A significant portion of the discussion focused on recent logical relations for session types that are indexed not only with a single type, but a set of types. These relations are necessitated by program equivalence statements such as noninterference, demanding observations along possibly several channels. The question came up whether such relations would also support proofs of free theorems in a linear setting based on a parametricity argument.

4.4 Real-World Applications of Behavioural Types

Kirstin Peters, Silvia Ghilezan, Jesper Bengtson, Christos Dimoulas, Marco Carbone, Felix Stutz, and Antonio Ravara

License © Creative Commons BY 4.0 International license
 © Kirstin Peters, Silvia Ghilezan, Jesper Bengtson, Christos Dimoulas, Marco Carbone, Felix Stutz, Antonio Ravara

- Why were attendees interested real-world applications?
 - Understanding how behavioural types can help with real-world projects and what is missing
 - Justifying need for research in grant proposals
- Some problems/examples for behavioural types
 - JEDIS bug (Java implementation of REDIS): was still trying to use a socket even though an exception was thrown before
 - tsunami prediction project
- More first-hand experiences
 - Racket contracts that only specify functional correctness and do not capture artificially introduced bugs in code; common issue that model is not expressive enough; need for specifying protocols
 - Erlang typing for `gen_server`: practically driven, challenges with failures, gradual typing
- Practical problems with evaluation
 - How to set up experiments?
 - What are target measures? How does one show that number of bugs reduce when using behavioural types?
 - Getting companies interested in pilot projects
- Industrial collaborations
 - often different interests than research: small steps are enough and there is little theoretical work needed to support this

- actual porting from theoretical work to practice is a lot of work and proves only worth it if something is found (risky for PhD projects)
- wish for push-bottom technology (e.g. Bedrock systems but fine with writing specifications)
- issue that universities seem to focus quite a bit on funding opportunities with such collaborations rather than research content (tone down probably?)
- Possibilities for case studies/collaborations with more applied domains, even in CS
 - it is likely that computing-heavy have software that suffers from issues BT solve
 - hospitals as part of universities: problem with data-privacy and training that is needed
 - IOT and intermittent computing

4.5 Typing Across Heterogeneous Components

Roland Kuhn, Philipp Haller, Sam Lindley, Vasco T. Vasconcelos, Simon Fowler, Alceste Scalas, Malte Viering, and Raymond Hu

License © Creative Commons BY 4.0 International license
 © Roland Kuhn, Philipp Haller, Sam Lindley, Vasco T. Vasconcelos, Simon Fowler, Alceste Scalas, Malte Viering, Raymond Hu

A large barrier to the widespread adoption of behavioural typing disciplines is that we must, at present, select a single tool and write our entire system using that tool. In practice at the very least we want to be able to design our system to make use of different components that use different, but not completely unrelated, behavioural type disciplines and tools (for example, having part of the system able to statically verify data constraints using refinement types, and another part of the system be able to check this using some form of monitoring).

There has been some work on runtime monitoring against local session types, but at present the results there only state that an invalid incoming message was dropped. There was a desire to go further than this, for example reporting failures, requesting message re-sends / reporting rejections to senders, or perhaps doing some message reordering.

Two related topics arose out of the discussion.

The first was a language design issue: what language features should our tools, at a minimum, support in order to allow interoperability? We thought that some way of handling failures (e.g., through exceptions) was probably important in this regard, although sometimes challenging to integrate into tools.

The second issue was about reconciling compatible yet subtly-different protocols, and a paper by Dezani et al. on session isomorphisms seemed important here. There was an extensive discussion about the notion of some sort of intermediate representation that could be used to implement adapters that can for example reorder messages and manipulate data sent along the wire in order to realise these transformations in practice in a uniform way.

We hope to write a position paper for PLACES in the coming weeks.

4.6 Probabilistic Behavioural Types

Emilio Tuosto, Silvia Ghilezan, Emanuele D’Osualdo, Jorge Pérez, Nobuko Yoshida, Marco Carbone, Marco Peressotti, Kirstin Peters, and Alceste Scalas

License © Creative Commons BY 4.0 International license
 © Emilio Tuosto, Silvia Ghilezan, Emanuele D’Osualdo, Jorge Pérez, Nobuko Yoshida, Marco Carbone, Marco Peressotti, Kirstin Peters, Alceste Scalas

Why probabilistic behavioural types?

Emilio: the answer is not obvious. My coauthors and I started working on them as a specification of a desired probabilistic process behaviour, that we use for monitoring <https://doi.org/10.1016/j.scico.2022.102847>, https://doi.org/10.1007/978-3-030-78142-2_7. I also have a work on probabilistic analysis of session types <https://doi.org/10.4230/LIPIcs.CONCUR.2020.14>.

Silvia Ghilezan: my interest comes from the connection between probabilistic lambda calculus and probabilistic logic

Emanuele D’Osualdo: my interest comes from my work on hyperproperties, I would like to look at systems that are probabilistic in nature

Jorge Perez: I am not sure I understand the meaning of probabilistic session types. Even if we have a probabilistic calculus (like e.g. probabilistic pi-calculus) what do we gain by putting probabilities in types?

Nobuko Yoshida: I am interested in understanding how to encode probabilistic lambda calculus in a session typed calculus, and it seems we need probabilistic (multiparty) session types

Marco Peressotti: My main motivation is understand how to extend choreographic programming to express and reason about randomness (e.g., randomness in algorithms, in the underlying communication model etc.) and eventually quantum protocols.

Kirstin Peters: I have both practical interest and quantum systems, which have probabilistic LTSs. We may use probabilistic session types to check that e.g. a program deadlocks with a certain probability.

Alceste Scalas: I share Jorge’s doubts (in fact I have only worked on probabilistic session types as specifications for for run-time onitoring, with Emilio et al). I have a bit of difficulty in seeing the “killer application” where putting probabilities in a session type gives you a clearly useful verification technique that may not be achieved using other analysis techniques for probabilistic programs.

Marco Carbone: I am also looking for applications that would clarify my understanding, like Jorge and Alceste Applications of probabilistic session types.

Emilio: Ugo dal Lago’s paper uses probabilistic session types for modelling ‘cryptographic experiments: <https://doi.org/10.4230/LIPIcs.CONCUR.2022.37>.

Marco Peressotti: Self-stabilising algorithms may also be a field of application

Marco Carbone: Tools like Prism allow for writing models with probabilities, but some models cannot be model-checked due to the state explosion problem. Could a type discipline like probabilistic session types help?

Nobuko Yoshida, Kirstin Peters, Marco Carbone, Marco Peressotti, Emilio Tuosto: we could look at probabilistic CCS examples, probabilistic dining philosophers, and see how probabilistic session types could allow e.g. to abstract and reduce the state space while still deriving useful probabilistic information (e.g. discarding low-probability events that may not relevantly impact the result). Probabilistic behavioural types could reduce the amount of concurrency to verify, and speed up the verification. Kirstin Peters has related work on the topic https://doi.org/10.1007/978-3-030-32505-3_12.

Jorge: on the topic of session-based execution optimisation, there is also related work by Luis Caires and Bernardo Toninho at ESOP 2024. Still, it is not clear whether bringing these optimisations in a probabilistic setting (e.g. to optimise analysis against PRISM model checking) would require probability annotations, maybe just having session-structured interaction is enough. We discuss a leader election protocol modelled in PRISM, and the properties that are verified (e.g. leader eventually elected with probability 1), and queries (i.e. computing the probability that a property eventually holds).

Everyone: the most popular probabilistic functional programming system should be Anglican <https://probprog.github.io/anglican/>. Recent work on quantum computing: <https://doi.org/10.1145/3632885>.

Emanuele: there are works proposing type systems for compositional abstraction for reasoning about the privacy of programs, with programs able to add noise to the computation (with some given probability annotations). eg <https://doi.org/10.1145/3009837.3009884>. We may do something similar for concurrency, e.g. prove that deadlocks may happen or not with some probability, depending on some probabilistic information.

Kirstin Peters: I see two lines: Adding probabilities to session types: do we gain anything from it? Take a program LTS with probabilities as input, but session types have no probabilities (a bit like probabilistic lambda calculus): how would it work?

Emilio: what if you put probabilities on the payload carried by messages? E.g. I may send an integer with probability p , or a string with probability $1-p$? Can we get something?


Jorge, Alceste: mention work by Padovani et al. on fair termination; uses non-deterministic choices, what if they are probabilistic? Emanuele believes that even in this case it may not be necessary to annotate types with probabilities (or probability intervals) to achieve probabilistic results, because e.g. the work on probabilistic privacy does not do it.

Conclusion

The conclusion of the breakout group is that there is no conclusion: there is no clear direction for having behavioural types with probabilities, so multiple attempts at exploring different directions are necessary. It is possible that specific problems may suggest solutions that naturally lead to probability-annotated session types, but that is not clear yet.

4.7 Open World Choreographies

Andrew Hirsch, Lukasz Ziarek, Marco Peressotti, Malte Viering, Raymond Hu, and Roland Kuhn

License  Creative Commons BY 4.0 International license
© Andrew Hirsch, Lukasz Ziarek, Marco Peressotti, Malte Viering, Raymond Hu, Roland Kuhn

What are “open” choreographies and possible directions

AH: started with a short introduction to the principles of choreographic programming and what makes the current state of the art “closed world”.

MP: Choreographies are global descriptions of the interactions of a distributed system but they assume that all participants in the system are fully described by the choreography.

AH: Focus on partially specified choreographies. Starting point to fix the discussion on a concrete example: DB-WS-Client(s) where a choreography specifies only DB and WS pushing assumptions on the Client(s) behaviour.

RK: Monitors at the boundary seem to be able to address several concerns raised in the example.

AH, LK: Suggested using some form of MPST to describe the interaction across the boundary. RK, MP: MPST are not expressive enough there is a need to identify a participant across multiple endpoints across the boundary.

MP: Asked to clarify the limitations of https://doi.org/10.1007/978-3-642-40184-8_30 or other existing works.

MV: A safe approach could be to synthesise “minimal” realisations for the unspecified part and use these to ensure properties.

RK: Example of “pluggable” choreographies used.

AH, MP: ChorLambda, Choral, and Pirouette provide higher-order composition of choreographies but still the end artefact requires full specification.

MP: Initial work on extension of ChorLambda and Choral with types with existential and universal quantification over roles. These allow to specify code for roles that join choreographies dynamically.

RH: Overall, either the choreographies allow for “step inside the boundary” or a monitor.

RH, MP: pointers to works that added compositionality to models initially “closed”. For instance compositional Petri Nets, Milner’s Bigraphs and IPOs, Graph rewriting DPOs.


Conclusions

From the discussion we identified to overall approaches:

1. Underspecified choreographies that identify participants without providing a full implementation of their behaviour in the choreography. The boundary between fully specified and underspecified requires some form of contract and runtime checks.
2. Choreographies provide mechanisms for adding new participants dynamically. In combination with support for cohorts of roles with uniform behaviour this can allow e.g. program systems with peers that can dynamically join or leave. An approach currently under exploration in Choral and ChorLambda are role quantifiers.

4.8 Mechanisation of Behavioural Types

Jesper, Luis, Kirstin, Robbert, Jonas, Alceste, Ralf

License  Creative Commons BY 4.0 International license
© Jesper, Luis, Kirstin, Robbert, Jonas, Alceste, Ralf

Jesper: defining a logic into a proof assistant is hard. Because of adequacy.

Again: depending on what you need to do, you need to pick the right proof assistant and techniques

Binders seem to be the main issue. Robert: avoid binders, but not clear how.

Discussion on nominal => Jesper thinks it’s by far the best way to deal with binders, if a deep embedding is the goal.

Luis: why hard in pi-calculus and easy in lambda. Coq can do it for you. In a functional language.

Ok, they are talking about embedding, deep vs shallow. This can make a difference.

- Deep embedding, hard.

- Shallow embedding.

==> Design a framework for working on session type mechanisation. Using Iris as an example.

Luis: the use of binders is different from pi to lambda. Substitution: names for names only.

Jesper & Kirstin: use nominal for deep embedding.

Robbert explains shallow vs deep. A shallow embedding allows to make the language modular (you can add new stuff).

Jesper: use nominal!

Kirstin: there is a need to reach out to people who are not expert.

Robbert: mechanise for confidence and mechanise for improving the proof assistant.

Jonas: having tutorial material.

Message: write about experiences, people can reuse it.

Ideally we should have a framework where to mechanise behavioural types (and linear stuff).

Robbert: If you want to build something more traditional, then somebody needs to make an effort, and perhaps nominal is a good way of doing it. There is also the Iris approach (shallow embedding) – this is great but (Kirstin says) it is hard to use (given the decoupling from syntax). (Kirstin says) An explanation of how the embedding allows to model Process calculi and about the advantages in proving in combination with a tutorial for how to get things working might allow also non-experts maybe without an understanding of the relation to logic to use this framework.

Robbert: third approach, intrinsic typing – (Agda?)

Alceste: Concurrency benchmark.

Ralf: The benchmark is excluding semantic approaches (like logical relations).

4.9 Dependent Session Types

Introduction round: what do we want out of the session?

- session types are important (send number N , followed by N messages where $N = 5$)
- systems: level-dependent session types, Actris
- refinement types: not allowing the highest bidder to bid again until a yet higher bid has been submitted
- what are dependent types in the context of session types? type families? type-level computation? indexed types?
- dependent types are functions from values to types, where values should be behaviours (which can encode arbitrary data)
- should you be able to depend on messages that you haven't observed?
- tracking dependency structures in multi-party sessions is non-trivial (it is intuitive in binary sessions, however)
- logistics auction: only the winning participant can continue the process after the auction

Example discussion:

- Simon: sending JSON, not labels
 - Actris can do this by using Coq “if/else” on the type-level
- Roland: factory logistics bidding
 - Brigitte proposes «Message-aware session types» (ESOP24) as a possible solution
 - importantly, the continuation may depend on what messages the process got from other channels prior
- Jonas: round-trip example ($A \rightarrow B \rightarrow C \rightarrow A$)

- bind variable x on the first send, then refer to it for the other sends to ensure that all send the same value
- the problem is that $B \rightarrow C$ has a binder (by duality) at B , but none at C
- solution might be to trace who will learn of the value and permit those roles to have binders when they receive it
- even more complicated is when only a location is sent around pointing to a linear resource
- the setting here specifically is that the global type is not known, the endpoint types are known and need to be combined to figure out whether the composition is well-typed (basically recovering a global type)

Participants

- Sören Auer
TIB – Hannover, DE
- Piero Andrea Bonatti
University of Naples, IT
- Juan Cano de Benito
Technical University of Madrid, ES
- Andrea Cimmino
Polytechnic University of Madrid, ES
- Michael Cochez
VU Amsterdam, NL
- John Domingue
The Open University – Milton Keynes, GB
- Michel Dumontier
Maastricht University, NL
- Nicoletta Fornara
University of Lugano, CH
- Irimi Fundulaki
FORTH – Heraklion, GR
- Sandra Geisler
RWTH Aachen, DE
- Anna Lisa Gentile
IBM Almaden Center – San Jose, US
- Paul Groth
University of Amsterdam, NL
- Peter Haase
Metaphacts GmbH – Walldorf, DE
- Andreas Harth
Fraunhofer IIS – Nürnberg, DE
- Olaf Hartig
Linköping University, SE
- James A. Hendler
Rensselaer Polytechnic Institute – Troy, US
- Aidan Hogan
University of Chile – Santiago de Chile, CL
- Katja Hose
TU Wien, AT
- Luis-Daniel Ibáñez
University of Southampton, GB
- Ryutaro Ichise
Tokyo Inst. of Technology, JP
- Ernesto Jiménez-Ruiz
City – University of London, GB
- Timotheus Kampik
Umeå University, SE & SAP Berlin, DE
- Sabrina Kirrane
Wirtschaftsuniversität Wien, AT
- Manolis Koubarakis
University of Athens, GR
- Luis C. Lamb
Boeing Research & Technology – Seattle, US
- Julian Padget
University of Bath, GB
- Harshvardhan J. Pandit
Dublin City University, IE
- Heiko Paulheim
Universität Mannheim, DE
- Axel Polleres
Wirtschaftsuniversität Wien, AT
- Philipp D. Rohde
TIB – Hannover, DE
- Daniel Schwabe
Rio de Janeiro, BR
- Oshani Seneviratne
Rensselaer Polytechnic – Troy, US
- Elena Simperl
King’s College London, GB
- Chang Sun
Maastricht University, NL
- Aisling Third
The Open University – Milton Keynes, GB
- Ruben Verborgh
Ghent University, BE
- Maria-Esther Vidal
TIB – Hannover, DE
- Sonja Zillner
Siemens AG – München, DE



Reviewer No. 2: Old and New Problems in Peer Review

Iryna Gurevych^{*1}, Anna Rogers^{*2}, Nihar B. Shah^{*3}, and
Jingyan Wang^{†4}

1 Department of Computer Science, TU Darmstadt, DE.
iryna.gurevych@tu-darmstadt.de

2 Department of Computer Science, IT University of Copenhagen, DK.
aarog@itu.dk

3 Machine Learning and Computer Science Departments, Carnegie Mellon
University – Pittsburgh, US. nihars@cs.cmu.edu

4 Georgia Institute of Technology – Atlanta, US. jingyanw@gatech.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 24052 “Reviewer No. 2: Old and New Problems in Peer Review”. This seminar provided a point of reflection on decades of personal experience of the participants in organizing different kinds of peer-reviewed venues in Natural Language Processing (NLP) and beyond, enabling an in-depth discussion of what has been tried, what seems to work and what doesn’t. The outcomes of the seminar include a white paper co-authored by most of the seminar participants, which outlines the research program, methodological and empirical challenges for NLP for peer review. The discussions at the seminar also resulted in several concrete policy proposals and initiatives, some of which are already in motion at the Association for Computational Linguistics and elsewhere.

Seminar January 28 – February 2, 2024 – <https://www.dagstuhl.de/24052>

2012 ACM Subject Classification Computing methodologies → Natural language processing;
Information systems → Expert systems; Information systems → Web applications

Keywords and phrases Peer Review, Natural Language Processing

Digital Object Identifier 10.4230/DagRep.14.1.130

1 Executive Summary

Anna Rogers (IT University of Copenhagen, Denmark, aarog@itu.dk)

Nihar Shah (Carnegie Mellon University, USA, nihars@cs.cmu.edu)

Iryna Gurevych (TU Darmstadt, Germany, iryna.gurevych@tu-darmstadt.de)

License  Creative Commons BY 4.0 International license
© Anna Rogers, Nihar Shah, and Iryna Gurevych

Background

Peer review is the best mechanism for assessing scientific validity of new research that we have so far. But this mechanism has many well-known issues, such as the different incentives of the authors and reviewers, difficulties with preserving reviewer and author anonymity to avoid social biases [22, 58, 68, 50, 39], confirmation and other cognitive biases [71, 16, 1, 32, 64], that even researchers fall prey to. These intrinsic problems are exacerbated in interdisciplinary fields like Natural Language Processing (NLP), where groups of researchers may vary so much in their methodology, terminology, and research agendas, that sometimes they have trouble even recognizing each other’s contributions as “research” [53].

* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed
under a Creative Commons BY 4.0 International license

Reviewer No. 2: Old and New Problems in Peer Review, *Dagstuhl Reports*, Vol. 14, Issue 1, pp. 130–161

Editors: Iryna Gurevych, Anna Rogers, and Nihar Shah



DAGSTUHL
REPORTS

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

Our Dagstuhl Seminar covered a range of topics related to organization of peer review in NLP, Machine Learning (ML), and venues more broadly in Artificial Intelligence for intelligent support of peer-reviewing, including the following:

- Improving the paper-reviewer matching by processes/algorithms that take into account both topic matches and reviewer interest in a given research question.
- Peer review vs methodological and demographic diversity in the field.
- Better practices for designing review forms and peer review policies.
- Improving the structural incentives for reviewers.
- Use of NLP and ML for intelligent peer reviewing support: increasing the quality and efficiency of peer review, opportunities and challenges.
- Peer-reviewing and research integrity.

Goals

We intended for the seminar to serve as a point of reflection on decades of personal experience of the participants in organizing different kinds of peer-reviewed venues in NLP and beyond, enabling an in-depth discussion of what has been tried, what seems to work and what doesn't. The objectives of the seminar included collaborative research on the methodological challenges of peer review, NLP and ML for intelligent support of peer-reviewing and actionable proposals, for example for paper-reviewer assignment policies and peer reviewing guidelines and workflows, informed by the experience of participants as chairs, editors, conference organizers, and reviewers.

Outcomes

The seminar was attended by researchers at different levels of seniority and from a variety of research backgrounds. While a large number of the attendees represented the Natural Language Processing community, about a third represented other communities within the broader sphere of Machine Learning. Most discussions focused on the peer review in the world of ultra-large conferences with thousands of submissions, but we also had a senior representative from fields where journals are most prominent, and hence an opportunity to learn from their experience.

Knowledge Sharing

The seminar started by contributed talks by a diverse group of participants (see section 3), which allowed us to share relevant experience and research findings pertinent to the topics of the seminar, across communities. Peer review issues are at most discussed in the business meetings of specific conferences, and there are hardly any opportunities to share this knowledge across communities. Hence, this knowledge-sharing section of the seminar by itself has been unique, and it proved to be useful to establish a common ground and points of reference for subsequent work during the seminar.

Problem elucidation

After the contributed talks, all the subsequent work was organized into breakout sessions (two running in parallel) on the following topics:

- Integrity issues in peer review (2 sessions)
- Diversity issues in peer review (3 sessions)

- Assisting peer review with NLP (3 sessions)
- Peer review policies (2 sessions)
- Incentives in peer review (3 sessions)
- Paper-reviewer matching (3 sessions)

The work in all these sessions combined brainstorming, establishing common ground and terms, discussing practical solutions for specific problems that were tried in various communities represented by the participants, and ideas for the future. Summaries of work in all the above topics are provided in section 4.

There were also two slots reserved for unstructured breakouts, and every day concluded with an overall summary session in which the leads for various topics summarized the discussions in that day.

Research program and community formation

The key outcome of the seminar is a white paper with the working title “*What Can NLP do for Peer Review?*”, co-authored by the majority of the participants of the seminar. It formulates the goals and research agenda of assisting peer review with NLP techniques, and we hope that it would play a key role in shaping this research field. This paper is available at [80]. It is accompanied by a repository for tracking research papers in this area, available at <https://github.com/OAfzal/nlp-for-peer-review>.

Concrete policies

The work in various breakout sessions culminated in the proposal of a new peer review committee for the Association of Computational Linguistics (ACL), that would oversee the systematic research and data-driven peer review policy development in the NLP community. This proposal has already been formally submitted to the ACL board, and generally approved. The work on formally establishing and announcing the committee will be finished in 2024.

Research problems and collaborations

This Dagstuhl Seminar also helped surface and crystallize a number of open problems, and alongside, helped establish inter-disciplinary collaborations for working on them, which may not have happened if not for this seminar.

Next steps

This Dagstuhl Seminar brought together an international, community of NLP and ML researchers from academia and industry to discuss the problems with peer review in large-scale conferences. This is a topic for which various subcommunities have different practices, expectations, and strong opinions, and the seminar brought much discussion throughout all days of the seminar (and also long into the night). This was also a unique opportunity to share the lessons learned the hard way, on issues which are often misconstrued as merely organizational issues. In fact, this is something to be seriously discussed as a research problem, for which much conceptual and empirical work is needed.

We hope that this seminar was the first in a series of events devoted to this topic, and that this inaugural event proves pivotal in the formation of a cohesive research community. The white paper prepared as the main outcome of this seminar aims to galvanize the NLP

and ML communities by offering them a wide selection of realistic research problems with peer review as an application area.

2 Table of Contents

Executive Summary

<i>Anna Rogers, Nihar Shah, and Iryna Gurevych</i>	130
--	-----

Overview of Talks

Natural Language Processing Meets Scientific Argumentation: The Case of Peer Reviewing <i>Anne Lauscher</i>	136
Peer review as text-based collaboration <i>Iliia Kuznetsov</i>	136
Peer review at ACL'23 <i>Anna Rogers</i>	136
Mitigating Biases in Peer Review <i>Jingyan Wang</i>	137
Technical Pitfalls and Possibilities in a [Rolling] Review System <i>Jonathan Kummerfeld</i>	137
Better Peer Review via AI <i>Kevin Leyton-Brown</i>	137
HotCRP for Dagstuhl <i>Eddie Kohler</i>	138
Natural Language Processing for Peer Review Assistance <i>Nils Dycke</i>	138
Studies on Citation Influence and Prediction <i>Xiaodan Zhu</i>	138
Semantic Scholar & Peer Review <i>Tom Hope</i>	139
Some experiments in reviewing <i>Nihar B. Shah</i>	139
Using ARR to Tackle Climate Change <i>Roy Schwartz</i>	139
Evaluating the peer review process <i>Alexander Goldberg</i>	140
Optimization of Scoring Rules <i>Jason Hartline</i>	140
ARR Reflexions on 1.5 years <i>Thamar Solorio</i>	140
Working conditions and satisfaction of early career NLP researchers in the era of LLMs <i>Sheng Lu</i>	141
Ethics Reviewing in NLP <i>Margot Mieskes</i>	141

Working Groups

Working Group on Policies for Peer Review
Anna Rogers 142

Working Group on Diversity in Peer Review
Anna Rogers 144

Working Group on Paper-Reviewer Matching
Anna Rogers 147

Working Group on Incentives in Peer Review
Nihar Shah 149

Working Group on Natural Language Processing (NLP) for Peer Review
Nihar Shah 151

Working Group on Integrity in Peer Review
Nihar Shah, Iryna Gurevych 153

Participants 161

3 Overview of Talks

The ordering of the talks is randomized (as opposed to ordering alphabetically).

3.1 Natural Language Processing Meets Scientific Argumentation: The Case of Peer Reviewing


Anne Lauscher (Universität Hamburg, DE)

License  Creative Commons BY 4.0 International license
© Anne Lauscher

Peer reviewing is a prime example of scientific argumentation: the authors present their work, a scientific claim in the form of a scientific publication, to the reviewers. The reviewers then engage in a debate, arguing why the claim should or should not be accepted to the body of tentatively accepted knowledge in a field. In this talk, I argue that theories and approaches rooted in argumentation theory and NLP/ computational argumentation can thus be leveraged to effectively support the different actors within this process. For instance, I discuss the case of automatic rebuttal template generation based on Jui-Jitsu argumentation [47] – a theory that has been proposed for the case of anti-science argumentation.

3.2 Peer review as text-based collaboration

Iliia Kuznetsov (TU Darmstadt, DE)

License  Creative Commons BY 4.0 International license
© Iliia Kuznetsov

Text-based collaboration is at the core of modern knowledge work. Peer review is a prime example of text-based collaboration, where authors, reviewers and area chairs work together to improve the initial paper draft via review and feedback. I introduce InterText [30] – a major project at UKP lab dedicated to the modeling of text as a living object in context, which we instantiate in the domain of academic peer review. I will describe our graph-based document representation and three novel intertextual modeling tasks – pragmatic tagging, linking and versioning. I will present F1000RD – the first corpus for intertextual NLP, and will discuss the results of our annotation studies, analysis, as well as existing challenges and ways forward towards NLP for living texts in context.

3.3 Peer review at ACL'23

Anna Rogers (IT University of Copenhagen, DK)

License  Creative Commons BY 4.0 International license
© Anna Rogers

ACL'23 implemented several changes to the standard ACL peer review process, followed up with survey-based evaluations from reviewers and area chairs. This talk describes the most successful innovations, including (a) paper-reviewer matching based on area + contribution + language 3-dimensional criteria, (b) soundness+excitement scores replacing the single

“overall recommendation” score, (c) structured reviewer complaints and issue flagging, (d) new format for reporting on peer review data as part of conference proceedings, (d) updated reviewer guidelines & first ACL policy on generative AI.

3.4 Mitigating Biases in Peer Review

Jingyan Wang (Georgia Institute of Technology – Atlanta, US)

License © Creative Commons BY 4.0 International license
© Jingyan Wang

I describe a particular type of bias in peer review where authors provide ratings to the reviewers’ review quality [74]. In this setting, a particular type of bias is induced by the author’s outcome (i.e., whether the author’s paper is accepted or not). In this work, we propose mild ordering assumptions to model the bias, and design a debiasing algorithm to correct student ratings adaptively to the amount of bias vs noise in the data. I also briefly describe other types of human bias in ratings, including miscalibration [73] and different causes of sequential effects [72].

3.5 Technical Pitfalls and Possibilities in a [Rolling] Review System

Jonathan Kummerfeld (The University of Sydney, AU)

License © Creative Commons BY 4.0 International license
© Jonathan Kummerfeld

A rolling review system differs from conferences because it has memory across time. That introduces a range of challenges (i.e., “pitfalls”) as well as opportunities (i.e., “possibilities”). I provide an overview of the technical infrastructure that is used by the ACL Rolling Review team, and discuss our experiences in the past and hopes for the future. In particular, I describe how tools can help with many reviewing tasks, but there is typically still a social dimension that is not solved. This context can help inform discussions in the seminar in terms of what exists today and what is possible tomorrow.

3.6 Better Peer Review via AI

Kevin Leyton-Brown (University of British Columbia – Vancouver, CA)

License © Creative Commons BY 4.0 International license
© Kevin Leyton-Brown

The talk summarized the reviewer-paper matching system used by Kevin and Mausam at AAAI 2021. The system decomposed into (1) collecting and processing input data; (2) formulating an optimization problem; (3) solving that problem; (4) two-phase reviewing. It also summarized an empirical analysis of data from the conference. The slides, but not the oral presentation, also briefly summarize a similar system for peer grading in large classes.

3.7 HotCRP for Dagstuhl

Eddie Kohler (Harvard University – Allston, US)

License  Creative Commons BY 4.0 International license
© Eddie Kohler

HotCRP is an online submission and review system used broadly in the CS theory, systems, networking, architecture, and security communities. Every reviewing community has values, as does every reviewing system; HotCRP’s values include speed, smoothness, ease of use, and openness to PC members. The talk also highlights some thoughts from earlier peer review discussions in other communities, and issues experienced by review system designers.

3.8 Natural Language Processing for Peer Review Assistance

Nils Dycke (TU Darmstadt, DE)

License  Creative Commons BY 4.0 International license
© Nils Dycke

Research on natural language processing (NLP) to support reviewers, authors and editors in the academic peer review process is still in its early stages. NLP for peer reviewing assistance holds promise for improving the quality of reviewing and increasing the efficiency of the process. In this talk I give an overview of the hurdles faced in the early stages of this young research field including the scarcity of open data, the lack of practical tasks, and the need for tools to disseminate NLP models to support peer review. I explain the rationale behind why and how NLP can help reviewers improve their work. Finally, I highlight our past work addressing these challenges, encompassing our data collection at ARR [11], the peer reviewing data corpus NLPEER [12], and the reading assistance tool CARE [79].

3.9 Studies on Citation Influence and Prediction


Xiaodan Zhu (Queen’s University – Kingston, CA)

License  Creative Commons BY 4.0 International license
© Xiaodan Zhu

We believe that most papers are based on a set of essential references. By an essential reference, we mean a reference that was highly influential or inspirational for the core ideas of the citing paper. In this talk, I first describe our previous research on predicting influential references from non-influential ones. Then, I move on to present our recent work on citation prediction in the legal domain, where citations are a foundation for many legal decision-making processes. I specifically present a prototype-based model that has some built-in interpretability for legal citation prediction.

3.10 Semantic Scholar & Peer Review


Tom Hope (The Hebrew University of Jerusalem, IL)

License  Creative Commons BY 4.0 International license
© Tom Hope

I presented some work done by AI2’s Semantic Scholar group which can be useful for building NLP-powered peer review systems. This includes the Semantic Scholar API which allows developers to access rich publication and author information from the Semantic Scholar Academic Graph, the Bridger tool for author matching based on shared methods and tasks authors work on, the Aspire scientific document embedding model for finding related papers, and the Semantic Reader platform which can support enhanced interactive reading experiences. Finally, I presented some of our recent work directly focused on peer review, led by Mike D’Arcy and Doug Downey: ARIES, which focuses on matching review comments to edits and generating edits in response to comments [9], and MARG, which introduces a multi-agent system for automatically generating reviews [8].

3.11 Some experiments in reviewing

Nihar B. Shah (Carnegie Mellon University – Pittsburgh, US)

License  Creative Commons BY 4.0 International license
© Nihar B. Shah

We discuss a set of experiments in scientific reviewing:

1. **Preprinting:** An anonymous survey of reviewers in (dual anonymous) ICML and EC conferences on whether they searched for their assigned papers online, finding that over a third of the reviewers do so [50].
2. **Author perceptions:** An experiment in NeurIPS 2021 finding that authors significantly overestimate the chances of their own papers getting accepted, and that co-authors significantly disagree on the relative merits of their co-authored papers [49].
3. **Discussions:** A randomized controlled trial at UAI 2022 on whether to show reviewers each others’ identities or not [48].
4. **Rebuttals:** A randomized controlled trial which finds no evidence of reviewers anchoring to their original opinions [35].
5. **AI reviewing:** A “chimera” test which finds that AI reviewers are unable to call out a nonsensical paper formed by pasting together parts of multiple papers (unpublished); and an evaluation of LLMs on their (in)capability to perform certain tasks in the review process [57].

3.12 Using ARR to Tackle Climate Change

Roy Schwartz (The Hebrew University of Jerusalem, IL)

License  Creative Commons BY 4.0 International license
© Roy Schwartz

The environmental effect of AI has been mostly studied in the context of the carbon footprint of models, while far less attention has been devoted to the cost of conference air travel. In this talk we present experiments showing that this factor also has a substantial environmental

cost. To partly mitigate this cost, we propose to allow authors to choose where to present their accepted papers, by allowing ARR to make accept/reject decisions. Our assumption is that many of them would prefer shorter travel if given the option. Our experiments show that this proposal could lead to substantial reductions in carbon emissions.

3.13 Evaluating the peer review process


Alexander Goldberg (Carnegie Mellon University – Pittsburgh, US)

License  Creative Commons BY 4.0 International license
© Alexander Goldberg

Evaluating outcomes and risks in the peer review process often proves quite challenging. This talk covers research on two aspects of evaluating peer review – (1) assessing review quality and (2) understanding privacy risks associated with open and transparent peer review. In evaluation of review quality, we highlight two biases that can arise in particular a positive bias towards (uselessly) longer reviews and bias by authors towards positive reviews [19]. On privacy risks, we describe deanonymization risk arising from revealing public comments on papers [18].

3.14 Optimization of Scoring Rules

Jason Hartline (Northwestern University – Evanston, US)

License  Creative Commons BY 4.0 International license
© Jason Hartline

This paper introduces an objective for optimizing proper scoring rules. The objective is to maximize the increase in payoff of a forecaster who exerts a binary level of effort to refine a posterior belief from a prior belief. In this framework we characterize optimal scoring rules in simple settings, give efficient algorithms for computing optimal scoring rules in complex settings, and identify simple scoring rules that are approximately optimal. In comparison, standard scoring rules in theory and practice – for example the quadratic rule, scoring rules for the expectation, and scoring rules for multiple tasks that are averages of single-task scoring rules – can be very far from optimal.

These scoring rules are applied to the task of grading peer reviews against TA reviews of homework in advanced undergraduate courses. Here the classical scoring rules give little incentive for effort and this incentive is improved by optimal scoring rules.

3.15 ARR Reflexions on 1.5 years

Thamar Solorio (MBZUAI – Abu Dhabi, AE)

License  Creative Commons BY 4.0 International license
© Thamar Solorio

The ACL Rolling Review (ARR) initiative launched in 2021 as a centralized review system for our *CL conferences. I've been serving as a co-editor in chief for ARR for 1.5 years now. In this talk I will present an overview of ARR, a snapshot of a typical two months reviewing

cycle. Then, I'll highlight some of the major challenges we face when trying to fulfill the reviewing needs of the community, while simultaneously being responsive to the requests for changes in our already tight reviewing process.

3.16 Working conditions and satisfaction of early career NLP researchers in the era of LLMs

Sheng Lu (TU Darmstadt, DE)

License  Creative Commons BY 4.0 International license
© Sheng Lu

The rapid development of large language models (LLMs) has led to challenges such as a lack of rigor in evaluation, an overwhelming amount of literature, and potentially negative impact on researchers' well-being due to the fast pace and a growing publication pressure. In June 2023, the Ubiquitous Knowledge Processing (UKP) Lab at the Technical University of Darmstadt and the Chair for Statistics and Data Science in Social Sciences and the Humanities (SODA) at Ludwig-Maximilians-Universität in Munich conducted an online survey with over 700 early career NLP researchers worldwide to gain insights into their working conditions and satisfaction in the era of LLMs. Even though these survey data do not allow us to make generalizable inferences, they provide important hints about the current working conditions and satisfaction of early career researchers in the NLP community. It would be beneficial for further research to include a more diverse range of respondents holding different types of positions and residing in different regions.

3.17 Ethics Reviewing in NLP

Margot Mieskes (Hochschule Darmstadt, DE)



License  Creative Commons BY 4.0 International license
© Margot Mieskes

For a couple of years now Ethics reviewing has been established as a part of the regular reviewing process in the NLP domain. But experience shows that this is a very time-consuming and far from structured process. In my talk I will present experience and lessons learned from being Ethics Co-Chair for three major NLP conferences (EMNLP 2021, EMNLP 2022 and LREC-COLING 2024) which used different reviewing platforms and the general chairs and program chairs had different levels of experience. I will also present some suggestions on how to improve the process and how to ensure that authors support the Ethics reviewing as part of the scientific process, rather than opposing it.

4 Working Groups

4.1 Working Group on Policies for Peer Review

Anna Rogers (IT University of Copenhagen, DK, arog@itu.dk)

License  Creative Commons BY 4.0 International license
 Anna Rogers

This working group focused on defining the scope of the peer review policies, and considering concrete problems in the communities with which the participants were familiar, and which could be addressed via various policies.

The discussion opened by considering what even counts as a policy. The following definitions were proposed: (a) a shared set of values and beliefs that evolves over time, (b) a long-term commitment, (c) a way to ensure consistent behavior by chairs etc.

This working group had 3 meetings in total, covering numerous topics. The topics that provoked the most discussion and suggested action items are summarized in this section.

4.1.1 Award policies

Currently most conferences do little to encourage good reviewer behavior. The ACL conferences recently created a policy to increase the number of reviewers and chairs nominated for awards to 1-1.5%¹, but this is still not enough: the chance to get such an award is still relatively small and not worth extra effort on the part of the reviewer. One suggestion was that if the awards are given to as much as 50% reviewers, this would reverse the incentive structure: *not* getting the award would create a negative social signal.

The current award offered to the outstanding reviewers at *ACL conferences is either free conference attendance as a virtual participant, or a discount on the in-person attendance. This does not necessarily reflect the needs of the reviewers, some of whom come from wealthy industry labs and do not need the monetary incentive. A survey could be organized to establish what other kinds of incentives could be useful. Some other reward ideas proposed in the discussion included:

- Sharing reviewer history to ORCID, given that it's possible to create generic service records there (e.g. conference names, number of papers reviewed, reviewer awards);
- 10% conference discount to 50% reviewers? (PeerJ case)
- Sharing generic reviewer history with potential employers (e.g., lab leaders looking for PhD students/postdocs), as it provides a useful signal about reliability and commitment;
- Likewise, area chairs could find it useful in grant applications if the area and statistics of their work could be shared to showcase community leadership in their research area (e.g. “outstanding area chair for question answering track, ACL 2024”);
- Various certificates showing different levels of achievement (reviewer certificate, outstanding reviewer certificate);
- A “star” system for reviewers where people can gain a star for good reviewer behavior (e.g., number of reviews, on time, high-quality, detailed reviews, emergency reviews), and feedback can be given to reviewer for moving up in the star scoring.

¹ https://2023.aclweb.org/program/best_reviewers/

4.1.2 Review form design

Most conferences, with which the participants were familiar, use unstructured or semi-structured review forms, with questions like “summary”, “strengths” and “weaknesses”. An alternative is to have structured review forms where reviewers are asked to evaluate various aspects separately. The group discussed evidence from Elsevier studies on the use of structured review forms [38], which increase agreement between reviewers, and decrease the cognitive load (allowing them to comment on the aspects of the paper for which they have expertise). From the perspective of program chairs, structured review forms can allow the chairs to better understand which parts of the paper have been reviewed more reliably, and mitigate issues like commensuration bias [32, 43]. Such forms could include concrete questions relevant to a specific paper type: e.g. “Is the statistical analysis sound?”, with the answer option “I’m not an expert on this”.

Structured review forms allow for better control to remove subjective categories in the review where biases from author identities are easier to creep in. If reviewers focus on the technical correctness of the work, these aspects might alleviate the problem of non double blind reviewing.

4.1.3 Institutional memory

At present, most conferences are organized on one-off basis, with most program chairs not having access even to reviewer history from the previous editions of the same conference, much less across venues. However, each cycle generates much useful information (late reviews, low-quality reviews, outstanding reviews etc.), and simply being able to track and use this information would probably help a lot in increasing the review quality. But there is no structure in place to maintain and organize this information. ACL Rolling Review could theoretically perform this for NLP community, as it already performs some long-term data collection [11], but there needs to be a broad mandate of storing and using this information for the organizational purposes.

It is not clear what should happen to the “bad” reviewers: simply de-prioritizing them in assignment process is not a negative incentive, since in practice it just means less work for them. Conversely, the “good” reviewers should not be rewarded by simply having more work assigned to them.

The group also discussed the privacy vs equity issue: peer review data is highly confidential and should not be disclosed without reviewer consent, especially since authors are often tempted to publicly bash their reviewers. At the same time, in some cases the privacy considerations protect the wrongdoers. One possibility to introduce reviewer consent to this process is to have reviewer agreements include an optional checkbox for granting the authors the right to use reviews however they choose, including for public discussion.

Finally, the group discussed the possibility of having a single reviewer profile within a system with an institutional memory, such as ARR, that would list various items from reviewer history, so as to visualize their impact on the community and highlight the fact that their reviewing record *is* tracked. This profile could include the following information: how many reviews they did, for what tracks, how many were late, how many were emergency reviews, what were the outcomes for the reviewed papers, how many best paper nominations (and the outcomes for those papers), ratings distribution for this reviewer vs conference mean, length of reviews vs mean, number of discussions vs mean, any feedback notes from area chairs, number and issue types flagged by the authors.

4.1.4 Findings Policy

Most of the current *ACL conferences have the more prestigious main track publication (e.g. proceedings of ACL), and a less-prestigious *Findings* <https://2020.emnlp.org/blog/2020-04-19-findings-of-emnlp> publication with a higher acceptance rate (usually 30-40%). The current workflow for peer review through ACL Rolling Review is that acceptance decisions are decoupled from review process, and are done independently by senior area chairs, sometimes months after the reviews were finished. This is unsatisfactory for the authors, who of course want faster decisions and not just reviews.

The group discussed the possibility of having ACL Rolling Review provide at least *Findings* decisions, which can be made purely on the ground of the judgement of technical soundness of the paper. Once such a decision was made, a *Findings* publication is guaranteed. Then the authors can still commit the paper to be considered for the main track publication. If the paper was rejected from *Findings*, it needs to go through the revise-and-resubmit process.

4.1.5 Next Steps: Peer Review Committee

Based on the above discussion, the seminar participants developed a proposal to the ACL executive board to establish the ACL Peer Review Committee: a working group dedicated to the development of peer review policies across ACL venues. This group would consider and help to develop proposals relevant to the peer review process, which originate either from ACL venues or the community. It would also monitor the implementation of any proposals, and ensure that the venues implementing them would consistently report on the results of any changes.

The proposal was approved, and the committee will be formally established in 2024. This committee will have a broad mandate to analyze internal peer review data from ACL venues for the purposes of developing evidence-based policies and improving the peer review process (but not for independent research by the committee members). For the sake of transparency, any results of such analysis will be made available as public reports.

While this committee will serve only ACL community, it is a good test case, since it has a lot of major conferences and already possesses the infrastructure for shared organization and institutional memory between them (ACL Rolling Review). The successful practices from this initiative could be shared with other communities.

4.2 Working Group on Diversity in Peer Review

Anna Rogers (IT University of Copenhagen, DK, arog@itu.dk)

License  Creative Commons BY 4.0 International license
© Anna Rogers

This working group had two meetings, focusing on a wide range of topics, some of which overlapped with the discussions in the Policy and Incentives groups. We started by noting the issues with even defining “diversity”: in peer review it is often discussed in terms of geographic diversity and levels of seniority, but it can have many other facets, such as representation of various topics, subfields and languages.

4.2.1 Reviewer Pool Representativeness

The conferences often start recruiting from the reviewer lists from past conferences, which means that a biased sample could keep being reused. Extra reviewers could be brought in through the networks of the chairs, which could also contribute to the bias. In ACL'23, there was a vast imbalance between the number of submissions and reviews contributed by Chinese researchers [54].

Perhaps the process should be more often, with the reviewers recruited through an open call and self-nomination via a sign-up form. Community groups such as Widening NLP could also help. To better estimate the extent of the problem, a conference registration form could include a question about reviewing (e.g. *Have you published here in the past 10 years? Have you reviewed here? If not, why?*)

For the specific China under-representation issue, the recommendation is to make sure that there are enough area chairs who are from China and based in China, and to ask them to help recruit widely from their networks. It would also help to connect with the National Science Foundation, there is an identifier system that is widely used within China.

4.2.2 Preprinting and Double-Blind Review

ACL used to have an embargo on posting papers on ArXiv prior to submitting to ARR or ACL conferences; but the policy recently changed to lift all restrictions.² It is still considered an integrity violation to search for papers one is reviewing, although some reviewers might use it a check for plagiarism. This change of policy necessitates discussion of how we can protect double-blind review, given that single-blind review is known to be influenced by demographic features associated with authors, such as country of affiliation, lab affiliation, fame, seniority etc. [46, 68, 39, 22, 59]. Among the possible ways to remedy this situation the seminar participants discussed the following:

- Peer review platforms could try to automate the checks for plagiarism, undisclosed preprints, previous publications etc. Then the reviewers could be told that they do not need to look for this. However, it is likely that many would still deliberately deanonymize the submissions [50].
- During paper-reviewer matching, the reviewers who disclosed the knowledge of a submission could be de-prioritized. However, the reviewers would have to input this information for all submissions that they are qualified to review. This also provides an extra opportunity to collusion rings [70, 34, 26], whose members could pretend to recognize work outside of the ring to increase the chances of assignments to each others' papers.
- The scores from reviewers who recognized the submission could be visually distinguished in the review reports presented to chairs, to help them downweight such scores (as they are potentially unreliable).
- "Confidence" score is too ambiguous, it could be interpreted as confidence about impact, novelty etc. It should be worded as confidence in the assessment of the technical aspects of the paper. Maybe this could help the reviewers to calibrate their assessments better even in a single-blind situation.

² <https://www.aclweb.org/portal/content/report-acl-committee-anonymity-policy>

4.2.3 Equity of Workload

A related issue is equity: peer-review work needs to be shared equally between recruited people, but the current conferences have the problem of too many qualified people leaving the reviewer pool after going into industry jobs. At the same time, we always have a lot of junior people who do not have the incentives to go through reviewer training. The following ideas were discussed to try to counter this trend and to have more equitable review loads:

- As an incentive, first-time reviewers could have a reduced load to help them spend more time on individual papers.
- First-time reviewers could have dedicated recognition/awards, and priority in bidding/assignments.
- First-time reviewers could be offered mentoring [63], and given the option to nominate their own PI as the mentor.
- Mentoring option in reviewer invitations: usually the invitations only have accept/decline options. There could be an option to nominate someone you would mentor.
- Venues could mandate a review load for the authors of papers submitted to a venue, with the possibility for authors who are not qualified or are contributing in other ways to be excused. For each author on submission, a form could be provided to indicate their availability as reviewers, with some common pre-set options to excuse some authors (e.g. “too junior”, “on leave”, “collaborator from a different field”, etc.).
- For papers with more than ten authors, require more authors to review.

4.2.4 Equity of Dissemination

Social Media

A phenomenon that was recently discussed is the “science influencers”: the mentions of research papers by certain Twitter accounts result in much increased popularity and citation counts for these papers [75]. Only few papers get promoted this way, and they could be more diverse in terms of geographical and gender representation of the promoted authors. Is it possible for the venues to do more to promote their accepted papers?

A conference could systematically collect content for social media from the authors for promoting the paper after it is accepted. Some journals already do that. On one hand, authors are more likely to better present their own work (more details, more engaging), but on the other hand if they are choosing to not self promote on social media –would they be willing to create social media content for the conference/editor?

ARR experimented with a bot for posting their anonymous preprints, but it wasn’t very popular. Perhaps one pitfall was that it was one bot for all tracks. Perhaps more specialized bots per track, or hashtags could help to filter the automatically posted content, and then it would be more useful?

Conferences

Some interventions for increasing the popularity of work from, for instance, underrepresented communities could happen during conference:

- Oral sessions could have “spillover effects”: having even one famous author in the panel could make it more likely that more people would attend other talks in the same oral session.
- Bidding data could be used to estimate which papers are more likely to be very popular, and to ensure that they are not all crowded together in the program.

- Random promotion of poster talks: selecting a certain fraction of posters to give a spotlight presentation of their work during a social event, and studying the impact of that additional exposure.

The attendance to conference themselves is of course far from equitable, due to the unequal distribution of funding and visa restrictions. Hybrid conferences have not been successful [52], and many venues are going back to mostly-onsite, with virtual participants as second-class citizens. One option to incentivize virtual attendance by decoupling the virtual event from the on-site event (e.g., as done at ACM EC 2024), and by offering free registration to the former [21]. The participants who could not attend on-site due to visa denials should automatically be given the option to present in the next conference where they could attend in person.

The current big conferences with thousands of attendees are very intimidating by themselves. Hence another option could be to subdivide conferences in 500 max attendees groups, and live stream tutorials for cross-location interactions. This should help bring both the cost and CO₂ impact³ down, but probably ups the organizational burden. Perhaps this could be organized as a distributed event at an international hotel chain, that has locations in many cities.

4.2.5 Next steps

For the deanonymization issue, the group concluded with the recommendation for the new ACL peer review committee (see subsection 4.1) to start tracking the information about preprints and intended preprints at ARR, and to monitor them over time to see what effect the new policy has, as compared to the impact of preprints reported at ACL'23 [54].

For the issue with reviewer diversity, ARR has already made an effort to broaden reviewer pool via recruitment of the authors of papers published at ACL. The effect on reviewer pool diversity needs to be estimated, and further measures for recruiting reviewers from China need to be taken if necessary. Equity of workload will be addressed at ARR by introducing a compulsory review load for authors of submitted papers.

4.3 Working Group on Paper-Reviewer Matching

Anna Rogers (IT University of Copenhagen, DK, arog@itu.dk)

License  Creative Commons BY 4.0 International license
© Anna Rogers

Paper-reviewer matching is a key element that has a lot of impact on the overall quality of peer review at large-scale conferences, and it is extremely important to get it right [51]. This working group had three meetings during the seminar. The discussion focused on the strategies of paper-reviewer matching and the lessons learned from the experience of the participants (as chairs and reviewers).

4.3.1 Assignment Strategies

The assignment is usually formalized as a discrete optimization problem, given the considerations of maximum load, the data about the quality of the match between submission and candidate reviewers, constraints such as conflicts of interests, reviewer seniority, experience

³ <https://gist.github.com/jacobeisenstein/ae0e13e270f3b00c9c2046b52297d018>

etc., and sometimes also prevention of strategic or dishonest behavior [20, 66, 6, 62, 28, 26, 10]. The discussion focused on the experience of AAAI [33] and ACL [54], as well as the theory conferences organized on HotCRP platform [29]. It became clear that in practice each platform relied on a unique set of complex constraints, and tuning these algorithms takes a lot of effort and expertise – and the result is difficult to evaluate.

The AAAI approach [33] was particularly complex and required a lot of variables to be set by hand. The group discussed the possibility of trying to learn the optimal parameters for such a system, but there is not enough data for this.

4.3.2 Assignment Criteria

Determining what constraints should be used for the assignments is a crucial step. The most salient component is the affinity score between (also called ‘similarity score’) the candidate reviewer and the submission, which is typically computed on the basis of reviewer publication history [6, 76, 7, 44, 41]. A major challenge is that these techniques may fail to pick up the aspects of similarity that are actually relevant for the match (e.g., abstracts can be stylistically similar but dissimilar in research topics), and NLP techniques have much room for improvement here [65]. Accordingly, the reviewer, author and chair trust in such scores is currently low [67]. Assignments based on past research may also no longer be interesting for the reviewers.

Another issue is the lack of clarity about the goal of review. Should the reviewers be optimized so that they would be most qualified to evaluate the technical correctness/soundness of the submission, or its novelty, or clarity of writing, or reproducibility, or excitement/interest to the community? It is not necessarily the case that these criteria coincide and would produce the same best match. Still, conference reviewers are implicitly expected to perform all these different roles, even though they may not be equally qualified for all this. In journals editors can craft per-paper committees; how can we do that on scale in conferences?

Another major challenge in paper-reviewer matching is noise in reviewer data (e.g., due to name disambiguation issues, unmaintained scholar profiles).

4.3.3 Bidding

Bidding is a process commonly used to directly elicit reviewer preferences. While that allows the reviewers to pick the papers they would be the most interested in, it has integrity risks (see subsection 4.6), and is quite laborious, which is why usually people only bid on a few papers shown at the top of the list [5, 13, 40]. Furthermore, nobody would like to bid on papers that look badly written or overall unpromising, but they still need to be reviewed. Finally, people often bid on what they want to learn about, not necessarily what they have expertise on.

There is potential for improving the bidding process by imputing bids: showing the reviewers not the full set of submissions, but an imputed subset to bid on. This could be done based on affinity scores, and so as to remove various potential conflicts of interest. (A similar approach was taken in [77] for the problem of collusion rings; see [24] for an evaluation and some pros and cons of it.) To help with deanonymization, the reviewers could be asked if they follow social media a lot, and if they do – they could be only shown non-preprinted submissions.

4.3.4 Standard for Paper-Reviewer Scoring

A big problem for research on paper-reviewer matching is that each conference operates on its own format for all the data that is used as constraints in the optimization problem. Any candidate solution needs to be integrated into this specific system, and once that is done – there is no “ground truth”, so it is hard to tell which alternative is actually better [56].

It would stimulate research in this area if there was at least a unified interface for paper-reviewer matching algorithms, used by all major conference platforms. Then any new solution could be tested more easily.

4.3.5 Matching in Iterative Assignment Setting

In HotCRP conferences as well as more recently other conferences in AI [33], it is common to have a variable number of reviews, due to multiple rounds of review, or other reasons. There is an initial pass, with a smaller number of reviews, meant to quickly reject obvious rejects, or accept papers that are good with high confidence. Then the remaining papers are assigned more reviewers, and their goal is to consider what was not covered by the first two reviewers. Perhaps some automated analysis of initial reviews could be used to facilitate picking the new reviewers, and explaining to them why they were assigned.

4.3.6 Next steps

- *Standard interface*: for the research on paper-reviewer matching to gain more traction in the community and become a research problem rather than just a conference organization problem, we need to develop a standard interface for interacting with confidential paper-reviewer matching data, that would be supported by the major conference platforms (OpenReview, Softconf, HotCRP).
- *Imputing bids*: The integrity, quality, and overall user experience of the bidding process can be enhanced by assigning a specific set of papers for each reviewer to evaluate and input their bidding information, and imputing the rest from it.

4.4 Working Group on Incentives in Peer Review

Nihar Shah (Carnegie Mellon University, Pittsburgh, US, nihars@cs.cmu.edu)

License  Creative Commons BY 4.0 International license
© Nihar Shah

Participants in this working group discussed two types of incentives: incentives for reviews for providing (high quality) reviews, and incentives for authors to submit only high-quality work. We discuss these two types of incentives in the following two subsections.

Although we present them separately for clarity of exposition, the discussions also captured some relations between the two. For instance, higher quality of reviews may reduce the number of submissions, if authors realize that low-quality submissions have little chance of getting in. Conversely, if authors were to curtail the number of submissions, the load on reviewers would reduce, and the quality of review may go up.

4.4.1 Incentives for Reviewers

Participants in this working group first discussed the various current incentives for reviewing:

- Prestige service roles, e.g., being area chairs or senior area chairs.
- People can mention it on their CVs.
- Building an informal social credit with colleagues that will lead to invitations (e.g., seminars) and positive response to future service requests.

- Listing and acknowledging of reviewers in the proceedings.
- Reviewer awards.
- Conference policies forcing eligible authors to also sign up to review.
- Additional motivations include keeping up with the literature, gatekeeping or influencing directions of their field, and improving scientific report quality [42].

Participants also discussed reasons why reviewers may not want to review [42]:

- Time may be better spent elsewhere.
- Too stressful.
- Assigned papers are not interesting.
- No recognition for the work.

Based on these observations, number of potential directions towards addressing this problem were then proposed and discussed:

- Assign reviewers in a manner that area chairs know the reviewers professionally, so that there is more accountability from reviewers.
- Collect (and possibly make public) reviewer performance over time.
- Overcome challenges in measuring review quality [19].
- If measuring review quality is hard, at least incentivize other measurable desiderata like completing the reviews in time or signing up for reviewing.
- Nudge people appropriately, e.g., via personalized reminders with actual names and the link to the paper they agreed to review, or personalized thank you emails for their service.
- Develop incentive structures where people who provide good reviews will have to review less.

4.4.2 Incentives for Authors

As for the incentives for authors, a key challenge discussed is the high prestige often attributed solely to the act of publishing a paper. Additionally, various organizations and governments provide monetary and other forms of incentives for publishing an increased number of papers. It has been observed that *“introduction of incentives by a country is associated with an increase in submissions by the country; the relation is particularly strong between cash bonuses and submissions”* [14]. Therefore, there are both implicit and explicit pressures on authors to publish more frequently.

In response, some academic conferences have implemented measures to discourage excessive submissions. For example, the International Conference on Learning Representations (ICLR) publicly posts all submitted papers, including those that are rejected, along with their peer reviews. This transparency is intended to deter submissions of lower quality, as the public availability of reviews can dissuade authors from submitting subpar work. Furthermore, several conferences now require authors to include reviews from any prior rejections when resubmitting papers, which are then passed to the new reviewers. This practice aims to prevent repeated submissions of low-quality papers, as a previous rejection could negatively influence subsequent reviews [64]. Despite these measures, the effectiveness of such policies in reducing the number of submissions or improving submission quality has not been comprehensively documented or measured. Thus, developing reliable methods to assess the causal impact of these policies on submission behaviors remains a critical open issue.

4.4.3 Next steps

Concrete next steps comprise:

- Developing more fair and accurate ways of measuring review quality, for instance, overcoming the problems discovered by the experiments in [19].
- Developing economic models capturing differences between venues where the peer-review quality is perceived to be good versus venues where it is perceived to be poor.
- Designing protocols for better longitudinal compilation of reviewer performance.

4.4.4 Open problems

There are two primary types of incentives that need thoughtful design to support the peer-review process. First, reviewing is often a voluntary task, and creating incentives that promote high-quality reviews is a challenge. While some strategies focus on increasing the volume of reviews, such as mandating that eligible authors must participate in peer reviewing, these do not necessarily guarantee quality. More targeted approaches aim to enhance review quality, like awards for outstanding reviewers such as those at the NeurIPS conference, as well as more theoretical approaches using game theory [78, 60, 69]. However, these approaches face several hurdles, including the gap between theoretical assumptions and real-world scenarios, unclear effects of these policies on reviewer motivation, and difficulties in accurately assessing the quality of reviews [19].

The second type of incentive concerns the authors. With the high numbers of submissions to conferences, ensuring thorough and high-quality peer reviews is becoming increasingly difficult. For authors, the cost of submitting papers is relatively low: acceptance means inclusion in a prestigious conference; rejection has minimal consequences. This low-risk environment coupled with noise in the process encourages the submission of papers that may even be of unsuitable quality, which might still be accepted. Additionally, some institutions and governments reward researchers for having papers accepted at these conferences, further incentivizing high submission rates and compounding the challenges in the review process. It is thus crucial to explore incentive systems that motivate authors to submit only high-quality work. Initiatives like ICLR policy of making all submissions public can deter low-quality submissions by adding repercussions for rejection. Meanwhile, platforms like TMLR accept all competent and relevant submissions, which could diminish the prestige of mere acceptance. The effectiveness of these initiatives in maintaining high standards, however, remains to be evaluated.

4.5 Working Group on Natural Language Processing (NLP) for Peer Review

Nihar Shah (Carnegie Mellon University, Pittsburgh, US, nihars@cs.cmu.edu)

License © Creative Commons BY 4.0 International license
© Nihar Shah

4.5.1 Overview

Natural Language Processing (NLP) has significant potential to improve the peer review process. There are various problems in peer review for which current works rely primarily on numerical data such as ratings provided by reviewers. These include challenges like miscalibration [17, 55, 73], subjectivity [43], elicitation [45, 37], and author-identity bias [68, 3].

Although these works focus on the numeric assessments, some of the main components of peer review—submissions and feedback—are text-based, making NLP a fitting tool for analysis and improvement.

Experiments in peer review also focus on quantifiable outcomes such as ratings and acceptance decisions [31, 2, 61, 63, 50, 64, 35]. Despite the numeric focus, the textual nature of the data suggests that NLP can offer substantial contributions to these areas. While some efforts have been made [39, 15, 48], much potential remains for NLP to further address these challenges.

Moreover, NLP can help tackle issues that are currently unaddressed, such as identifying unsubstantiated criticisms in reviews or pinpointing deficiencies in papers. However, developing these NLP methods must also consider the safety and fairness of their applications, and how these methods are evaluated and measured. This area of research is gaining increasing popularity [12, 36, 30, 9, 27], and all these considerations will be explored in detail in the forthcoming paper titled “What Can Natural Language Processing Do for Peer Review?”[80] emerging from this Dagstuhl Seminar.

4.5.2 Whitepaper

This working group quickly converged to the understanding that:

- There is a huge opportunity for improving peer review via latest advancements in natural language processing.
- There are also as many challenges as doing it in a safe, fair, and accurate manner, as well as in evaluating the outcomes.
- It is thus important to convey this message to researchers in NLP, ML and related communities, and also make it as easy as possible to step into this research application domain.
- A suitable means of doing so is to write a position paper, accompanied with a repository containing various datasets pertaining to peer review.

With this motivation, the remainder of the sessions focused on planning, organizing, and beginning to write the position paper and compile the datasets. The paper touches upon the following topics:

- Background of the peer-review process.
- Assistance before the review process.
 - Preparing the manuscripts.
 - * Writing assistance for authors.
 - * Helping authors form metadata such as keywords and TL;DRs.
 - * Initial screening of manuscripts for basic checks.
 - Reviewer-paper matching.
 - * Computing similarities between reviewers and submitted papers.
 - * Finding conflicts of interest.
 - * Reducing strategic behavior.
- Assistance during the review process.
 - Evaluating certain aspects of the manuscript.
 - Helping write the review.
 - Discussions with authors and reviewers.
- Assistance after the review process.
 - Helping with the meta review.
 - Final decisions.

- Camera-ready submissions.
- Post-conference analysis.
- Data, privacy, and legal aspects.
- Measurements and experimentations.
- Ethics.

4.5.3 Next steps

The whitepaper manuscript is available at [80] and the associated repository is available at <https://github.com/OAfzal/nlp-for-peer-review>.

4.6 Working Group on Integrity in Peer Review

Nihar Shah (Carnegie Mellon University, Pittsburgh, US, nihars@cs.cmu.edu)

Iryna Gurevych (TU Darmstadt, DE, iryna.gurevych@tu-darmstadt.de)

License © Creative Commons BY 4.0 International license
© Nihar Shah, Iryna Gurevych

This working group focused on issues undermining the integrity of peer review processes. Initially, participants explored a range of challenges affecting peer review integrity (see, e.g., [57, Section 4]). As discussions progressed, the consensus emerged that collusion rings and AI tools to support the integrity of the peer review process represented the most critical issues. Consequently, one subgroup dedicated the rest of their session to this specific problem. Another subgroup worked on developing a roadmap for AI tools and protocols for collecting peer review data as an enabling factor for the envisaged tools.

4.6.1 Collusion Rings

The allure of being published in prestigious conferences can sometimes encourage unethical behaviors among participants. One concerning trend that has gained attention is the formation of collusion rings [70, 34]. In these scenarios, groups of researchers manipulate the peer review system to review each other's papers. They then provide favorable evaluations, often disregarding the true merit of the work.

In recent years, program chairs have devoted considerable time and effort to addressing the issue of collusion rings. Tackling this challenge is essential as it directly undermines the integrity and fairness of the peer review process. Alongside this, there is a growing concern over bullying and abuse of power within the academic community. Participants reported instances where senior researchers, including some area chairs at conferences, have pressured junior colleagues to engage in these unethical activities.

One approach to addressing this issue is through technical means: developing algorithmic methods designed to detect or prevent collusion rings. Much research has already been conducted in this area [26, 77, 33, 4, 33, 24, 25, 23]. Further discussions in the working group focused on several potential strategies:

- Introducing additional conditions, such as requiring more reviewer bids. While every intervention has associated costs, it is crucial to consider potential drawbacks. For instance, although the ARR system does not involve bidding, collusion could still occur. Currently, assignments are automated with the option for Area Editors (AEs) to make changes. However, this could lead to issues if an AE is compromised.

- Implementing policies to ensure that the same person does not repeatedly review another individual's papers over multiple years. While such interventions encourages diversity, their disadvantages must also be weighed carefully in terms of reducing expertise of assigned reviewers.
- Approaching the problem as a network issue, where more distant social links might serve as indicators of conflicts of interest. Further analyzing patterns of paper submissions and reviews, identifying discrepancies between poor-quality papers and positive reviews. Although not necessarily indicative of malicious collusion, such patterns could still pose significant problems.
- Conducting automated assessments of papers and, in cases of high disagreement between reviewers, assigning an additional reviewer. This approach must be handled carefully to avoid a high rate of false positives.
- Developing models to quantify the likelihood that certain behaviors are due to chance. This requires careful evaluation to ensure accuracy and effectiveness.

A second, human-centric approach was explored to address this problem, with several strategies proposed:

- *Whistleblower Support*: Participants shared knowledge about instances where collusion rings were exposed through shared communications within chat groups. This highlights the need for robust mechanisms that enable whistleblowers to safely report unethical behaviors.
- *Policy Development Board*: Conferences should establish a board dedicated to policy creation. This board would be responsible for developing protocols and guidelines to manage reports of misconduct effectively, including those of handling whistleblower reports.
- *Guidelines for Program Chairs*: Since program chairs of conferences change every year, it will be useful to develop guidelines for program chairs on what sorts of prevention measures are there for collusion rings.
- *Inter-Conference Data Sharing*: Some preventive measures may require data from multiple conferences to detect recurring patterns of misconduct. However, this poses legal and technical challenges, such as difficulties in transferring data between different conferences or iterations due to current platform limitations. Addressing these issues will be crucial for effective prevention.
- *Education and Training*: Implementing educational programs to inform researchers about unethical practices and their severe consequences can serve as a preventive measure. This approach aims to cultivate a culture of integrity and transparency within the academic community.

Next steps: Collusion rings may be addressed either by developing methods to detect them, or by preventing them (during the reviewer assignment phase itself). Although there is ongoing research on these questions [26, 33, 77, 4, 24, 23, 25], finding solutions involves significant trade-offs, and this problem largely remains open [24, 23]. Effective strategies to combat these unethical practices are crucial for maintaining the integrity of the peer-review process. There are a number of subsequent steps that are underway following the Dagstuhl Seminar. The first is to make more researchers aware of this problem and encouraging them to work on technical solutions using the tools at their disposal. For instance, NLP tools have not been used so far to combat against collusion rings, and we argue for doing so in the forthcoming paper on NLP for peer review (discussed in the NLP section of this report). We are also developing guidelines for program chairs and pushing along policies for some sharing

of data across conferences. We envisage further outcomes to emanate in due time based on the aforementioned directions conceived at the Dagstuhl Seminar.

4.6.2 AI Tools

A major bottleneck for developing AI tools to support the integrity of peer review is the lack of data for training and evaluation in this sensitive domain subject to privacy and data protection. Existing large scale data collections suffer from the low-response rate. Therefore, the group participants were interested in helping with the data collection.

This involved tasks such as creating an empirical protocol for data collection, achieving the ethical and legal clearance by writing a proposal for the ACL ethics committee, etc., defining what kind of data to extract, for what purposes and how to implement this technically. Further questions that were covered in the discussion were data ownership, distribution and management, allowing for retraction of consent, backward compatibility with the previous data collection workflow, cross-community data collection, incentivizing data donation and public outreach through blog posts, social media, etc.

At the time of writing, most of the discussed topics have already been addressed by the seminar participants. The result is documented in this website: <https://arr-data.aclweb.org>. The data collection protocol has been successfully implemented and the data collection is now ongoing within the ACL Rolling Review (ARR) platform.

References

- 1 David M. Allen and James W. Howell, editors. *Groupthink in Science: Greed, Pathological Altruism, Ideology, Competition, and Culture*. Springer International Publishing, Cham, 2020.
- 2 Alina Beygelzimer, Yann Dauphin, Percy Liang, and Jennifer Wortman Vaughan. The NeurIPS 2021 consistency experiment. <https://blog.neurips.cc/2021/12/08/the-neurips-2021-consistency-experiment/>, 2021. Online; accessed 18-April-2024.
- 3 Rebecca M Blank. The effects of double-blind versus single-blind reviewing: Experimental evidence from the american economic review. *The American Economic Review*, pages 1041–1067, 1991.
- 4 Niclas Boehmer, Robert Bredereck, and André Nichterlein. Combating collusion rings is hard but possible. *arXiv preprint arXiv:2112.08444*, 2021.
- 5 Guillaume Cabanac and Thomas Preuss. Capitalizing on order effects in the bids of peer-reviewed conferences to secure reviews by expert referees. *Journal of the Association for Information Science and Technology*, 64(2):405–415, 2013.
- 6 Laurent Charlin and Richard Zemel. The Toronto Paper Matching System: An automated paper-reviewer assignment system. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28, Atlanta, Georgia, USA, May 2013. Journal of Machine Learning Research Workshop and Conference Proceedings.
- 7 Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel S Weld. Specter: Document-level representation learning using citation-informed transformers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2270–2282, 2020.
- 8 Mike D’Arcy, Tom Hope, Larry Birnbaum, and Doug Downey. MARG: Multi-agent review generation for scientific papers. *arXiv preprint arXiv:2401.04259*, 2024.
- 9 Mike D’Arcy, Alexis Ross, Erin Bransom, Bailey Kuehl, Jonathan Bragg, Tom Hope, and Doug Downey. ARIES: A corpus of scientific paper edits made in response to peer reviews. *arXiv preprint arXiv:2306.12587*, 2023.

- 10 Komal Dhull, Steven Jecmen, Pravesh Kothari, and Nihar B Shah. Strategyproofing peer assessment via partitioning: The price in terms of evaluators’ expertise. In *HCOMP*, 2022.
- 11 Nils Dycke, Iliia Kuznetsov, and Iryna Gurevych. Yes-yes-yes: Proactive data collection for ACL rolling review and beyond. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 300–318, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- 12 Nils Dycke, Iliia Kuznetsov, and Iryna Gurevych. NLPeer: A unified resource for the computational study of peer review. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5049–5073, Toronto, Canada, July 2023. Association for Computational Linguistics.
- 13 Tanner Fiez, Nihar Shah, and Lillian Ratliff. A SUPER* algorithm to optimize paper bidding in peer review. In *Conference on Uncertainty in Artificial Intelligence*, pages 580–589. Proceedings of Machine Learning Research, 2020.
- 14 Chiara Franzoni, Giuseppe Scellato, and Paula Stephan. Changing incentives to publish. *Science*, 333(6043):702–703, 2011.
- 15 Yang Gao, Steffen Eger, Iliia Kuznetsov, Iryna Gurevych, and Yusuke Miyao. Does my rebuttal matter? Insights from a major NLP conference. In Jill Burstein, Christy Doran, and Tamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1274–1290, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- 16 J. A. Garcia, Rosa Rodriguez-Sánchez, and J. Fdez-Valdivia. Confirmatory bias in peer review. *Scientometrics*, 123(1):517–533, April 2020.
- 17 H. Ge, M. Welling, and Z. Ghahramani. A Bayesian model for calibrating conference review scores. Manuscript, 2013. Available online <http://mlg.eng.cam.ac.uk/hong/unpublished/nips-review-model.pdf> Last accessed: April 4, 2021.
- 18 Alexander Goldberg, Giulia Fanti, and Nihar B Shah. Batching of tasks by users of pseudonymous forums: Anonymity compromise and protection. In *ACM SIGMETRICS*, 2023.
- 19 Alexander Goldberg, Ivan Stelmakh, Kyunghyun Cho, Alice Oh, Alekh Agarwal, Danielle Belgrave, and Nihar B Shah. Peer reviews of peer reviews: A randomized controlled trial and other experiments. *arXiv preprint arXiv:2311.09497*, 2023.
- 20 Judy Goldsmith and Robert H. Sloan. The AI conference paper assignment problem. *WS-07-10:53–57*, 12 2007.
- 21 Jason Harline. The Dissemination Game: Incentives of In-Person vs Virtual Participation – Communications of the ACM, July 2023.
- 22 Jürgen Huber, Sabiou Inoua, Rudolf Kerschbamer, Christian König-Kersting, Stefan Palan, and Vernon L. Smith. Nobel and novice: Author prominence affects peer review. *Proceedings of the National Academy of Sciences*, 119(41):e2205779119, October 2022.
- 23 Steven Jecmen, Nihar B Shah, Fei Fang, and Leman Akoglu. On the detection of reviewer-author collusion rings from paper bidding. *arXiv preprint arXiv:2402.07860*, 2024.
- 24 Steven Jecmen, Nihar B Shah, Fei Fang, and Vincent Conitzer. Tradeoffs in preventing manipulation in paper bidding for reviewer assignment. *arXiv preprint arXiv:2207.11315*, 2022.
- 25 Steven Jecmen, Minji Yoon, Vincent Conitzer, Nihar B Shah, and Fei Fang. A dataset on malicious paper bidding in peer review. In *Proceedings of the ACM Web Conference 2023*, pages 3816–3826, 2023.

- 26 Steven Jecmen, Hanrui Zhang, Ryan Liu, Nihar Shah, Vincent Conitzer, and Fei Fang. Mitigating manipulation in peer review via randomized reviewer assignments. *Advances in Neural Information Processing Systems*, 33:12533–12545, 2020.
- 27 Neha Kennard, Tim O’Gorman, Rajarshi Das, Akshay Sharma, Chhandak Bagchi, Matthew Clinton, Pranay Kumar Yelugam, Hamed Zamani, and Andrew McCallum. DISAPERRE: A dataset for discourse structure in peer review discussions. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1234–1249, Seattle, United States, July 2022. Association for Computational Linguistics.
- 28 Ari Kobren, Barna Saha, and Andrew McCallum. Paper matching with local fairness constraints. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1247–1257, 2019.
- 29 Eddie Kohler. HotCRP conference management software, 2013.
- 30 Ilya Kuznetsov, Jan Buchmann, Max Eichler, and Iryna Gurevych. Revise and Resubmit: An intertextual model of text-based collaboration in peer review. *Computational Linguistics*, 48(4):949–986, 12 2022.
- 31 N. Lawrence and C. Cortes. The NIPS Experiment. <http://inverseprobability.com/2014/12/16/the-nips-experiment>, 2014. Online; accessed 17-April-2024.
- 32 Carole J Lee. Commensuration bias in peer review. *Philosophy of Science*, 82(5):1272–1283, 2015.
- 33 Kevin Leyton-Brown, Yatin Nandwani, Hedayat Zarkoob, Chris Cameron, Neil Newman, Dinesh Raghu, et al. Matching papers and reviewers at large conferences. *Artificial Intelligence*, 2024.
- 34 Michael L Littman. Collusion rings threaten the integrity of computer science research. *Communications of the ACM*, 64(6):43–44, 2021.
- 35 Ryan Liu, Steven Jecmen, Vincent Conitzer, Fei Fang, and Nihar B Shah. Testing for reviewer anchoring in peer review: A randomized controlled trial. *arXiv preprint arXiv:2307.05443*, 2023.
- 36 Ryan Liu and Nihar B. Shah. ReviewerGPT? An exploratory study on using large language models for paper reviewing. *arXiv preprint arXiv:2306.00622*, 2023.
- 37 Yusha Liu, Yichong Xu, Nihar B Shah, and Aarti Singh. Integrating rankings into quantized scores in peer review. *arXiv preprint arXiv:2204.03505*, 2022.
- 38 Mario Malički and Bahar Mehmani. Structured peer review: Pilot results from 23 Elsevier journals. *bioRxiv preprint bioRxiv:10.1101/2024.02.01.578440*, February 2024.
- 39 Emaad Manzoor and Nihar B Shah. Uncovering latent biases in text: Method and application to peer review. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4767–4775, 2021.
- 40 Reshef Meir, Jérôme Lang, Julien Lesca, Nicholas Mattei, and Natan Kaminsky. A market-inspired bidding scheme for peer review paper assignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4776–4784, 2021.
- 41 Sheshera Mysore, Mahmood Jasim, Andrew McCallum, and Hamed Zamani. Editable user profiles for controllable text recommendation. *arXiv preprint arXiv:2304.04250*, 2023.
- 42 Syavash Nobarany, Kellogg S Booth, and Gary Hsieh. What motivates people to review articles? the case of the human-computer interaction community. *Journal of the Association for Information Science and Technology*, 67(6):1358–1371, 2016.
- 43 Ritesh Noothigattu, Nihar Shah, and Ariel Procaccia. Loss functions, axioms, and peer review. *Journal of Artificial Intelligence Research*, 70:1481–1515, 2021.
- 44 Justin Payan and Yair Zick. I will have order! Optimizing orders for fair reviewer assignment. *arXiv preprint arXiv:2108.02126*, 2021.

- 45 Michael Pearce and Elena A Erosheva. A unified statistical learning model for rankings and scores with application to grant panel review. *arXiv preprint arXiv:2201.02539*, 2022.
- 46 Douglas P Peters and Stephen J Ceci. Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, 5(2):187–195, 1982.
- 47 Sukannya Purkayastha, Anne Lauscher, and Iryna Gurevych. Exploring jiu-jitsu argumentation for writing peer review rebuttals. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 14479–14495. Association for Computational Linguistics, 2023.
- 48 Charvi Rastogi, Xiangchen Song, Zhijing Jin, Ivan Stelmakh, Hal Daumé III, Kun Zhang, and Nihar B Shah. A randomized controlled trial on anonymizing reviewers to each other in peer review discussions. *arXiv preprint arXiv:2403.01015*, 2024.
- 49 Charvi Rastogi, Ivan Stelmakh, Alina Beygelzimer, Yann N Dauphin, Percy Liang, Jennifer Wortman Vaughan, Zhenyu Xue, Hal Daumé III, Emma Pierson, and Nihar B Shah. How do authors’ perceptions of their papers compare with co-authors’ perceptions and peer-review decisions? *arXiv preprint arXiv:2211.12966*, 2022.
- 50 Charvi Rastogi, Ivan Stelmakh, Xinwei Shen, Marina Meila, Federico Echenique, Shuchi Chawla, and Nihar Shah. To ArXiv or not to ArXiv: A study quantifying pros and cons of posting preprints online. *arXiv preprint arXiv:2203.17259*, 2022.
- 51 Marko A Rodriguez, Johan Bollen, and Herbert Van de Sompel. Mapping the bid behavior of conference referees. *Journal of Informetrics*, 1(1):68–82, 2007.
- 52 Anna Rogers. Field Notes on Hybrid Conferences (EMNLP 2021), November 2021.
- 53 Anna Rogers and Isabelle Augenstein. What can we do to improve peer review in NLP? In Trevor Cohn, Yulan He, and Yang Liu, editors, *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1256–1262. Association for Computational Linguistics, November 2020.
- 54 Anna Rogers, Marzena Karpinska, Jordan Boyd-Graber, and Naoaki Okazaki. Program chairs’ report on peer review at ACL 2023. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 40–75, Toronto, Canada, July 2023. Association for Computational Linguistics.
- 55 Magnus Roos, Jörg Rothe, Joachim Rudolph, Björn Scheuermann, and Dietrich Stoyan. A statistical approach to calibrating the scores of biased reviewers: The linear vs. the nonlinear model. In *Multidisciplinary Workshop on Advances in Preference Handling*, 2012.
- 56 Martin Saveski, Steven Jecmen, Nihar Shah, and Johan Ugander. Counterfactual evaluation of peer-review assignment policies. In A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 58765–58786. Curran Associates, Inc., 2023.
- 57 Nihar B Shah. Challenges, experiments, and computational solutions in peer review. *Communications of the ACM*, 65(6):76–87, 2022.
- 58 Inna Smirnova, Daniel M. Romero, and Misha Teplitskiy. The bias-reducing effect of voluntary anonymization of authors’ identities: Evidence from peer review, January 2023.
- 59 Inna Smirnova, Daniel M Romero, and Misha Teplitskiy. The bias-reducing effect of voluntary anonymization of authors’ identities: Evidence from peer review. *Available at SSRN 4190623*, 2023.
- 60 Siddarth Srinivasan and Jamie Morgenstern. Auctions and prediction markets for scientific peer review. *arXiv preprint arXiv:2109.00923*, 2021.
- 61 Ivan Stelmakh, Charvi Rastogi, Nihar B Shah, Aarti Singh, and Hal Daumé III. A large scale randomized controlled trial on herding in peer-review discussions. *arXiv preprint arXiv:2011.15083*, 2020.

- 62 Ivan Stelmakh, Nihar Shah, and Aarti Singh. PeerReview4All: Fair and accurate reviewer assignment in peer review. *Journal of Machine Learning Research*, 22(163):1–66, 2021.
- 63 Ivan Stelmakh, Nihar B Shah, Aarti Singh, and Hal Daumé III. A novice-reviewer experiment to address scarcity of qualified reviewers in large conferences. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4785–4793, 2021.
- 64 Ivan Stelmakh, Nihar B Shah, Aarti Singh, and Hal Daumé III. Prior and prejudice: The novice reviewers’ bias against resubmissions in conference peer review. volume 5, pages 1–17. ACM New York, NY, USA, 2021.
- 65 Ivan Stelmakh, John Wieting, Graham Neubig, and Nihar B. Shah. A gold standard dataset for the reviewer assignment problem. *arXiv preprint arXiv:2303.16750*, 2023.
- 66 Camillo J Taylor. On the optimal assignment of conference papers to reviewers. 2008.
- 67 Terne Thorn Jakobsen and Anna Rogers. What factors should paper-reviewer assignments rely on? community perspectives on issues and ideals in conference peer-review. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4810–4823, Seattle, United States, July 2022. Association for Computational Linguistics.
- 68 Andrew Tomkins, Min Zhang, and William D. Heavlin. Reviewer bias in single- versus double-blind peer review. *Proceedings of the National Academy of Sciences*, 114(48):12708–12713, 2017.
- 69 Alexander Ugarov. Peer prediction for peer review: Designing a marketplace for ideas. *arXiv:2303.16855*, 2023.
- 70 T. N. Vijaykumar. Potential organized fraud in ACM/IEEE computer architecture conferences. <https://medium.com/@tnvijayk/potential-organized-fraud-in-acm-ieee-computer-architecture-conferences-ccd61169370d>, 2020. Online; accessed 17-April-2024.
- 71 Jian Wang, Reinhilde Veugelers, and Paula Stephan. Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, 46(8):1416–1436, October 2017.
- 72 Jingyan Wang and Ashwin Pananjady. Modeling and correcting bias in sequential evaluation. In *Conference on Economics and Computation, (EC)*, 2023.
- 73 Jingyan Wang and Nihar B Shah. Your 2 is my 1, your 3 is my 9: Handling arbitrary miscalibrations in ratings. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pages 864–872, 2019.
- 74 Jingyan Wang, Ivan Stelmakh, Yuting Wei, and Nihar Shah. Debiasing evaluations that are biased by evaluations. In *AAAI Conference on Artificial Intelligence*, 2021.
- 75 Iain Xie Weissburg, Mehira Arora, Xinyi Wang, Liangming Pan, and William Yang Wang. Tweets to Citations: Unveiling the Impact of Social Media Influencers on AI Research Visibility, March 2024.
- 76 John Wieting, Kevin Gimpel, Graham Neubig, and Taylor Berg-Kirkpatrick. Simple and effective paraphrastic similarity from parallel translations. In *ACL*, pages 4602–4608, Florence, Italy, July 2019.
- 77 Ruihan Wu, Chuan Guo, Felix Wu, Rahul Kidambi, Laurens van der Maaten, and Kilian Weinberger. Making paper reviewing robust to bid manipulation attacks. In *ICML*, 2021.
- 78 Yuanzhang Xiao, Florian Dörfler, and Mihaela Van Der Schaar. Incentive design in peer review: Rating and repeated endogenous matching. *IEEE Transactions on Network Science and Engineering*, 6(4):898–908, 2018.
- 79 Dennis Zyska, Nils Dycke, Jan Buchmann, Ilija Kuznetsov, and Iryna Gurevych. CARE: Collaborative AI-assisted reading environment. In Danushka Bollegala, Ruihong Huang, and Alan Ritter, editors, *Proceedings of the 61st Annual Meeting of the Association for*

Computational Linguistics (Volume 3: System Demonstrations), pages 291–303, Toronto, Canada, July 2023. Association for Computational Linguistics.

- 80 Iliia Kuznetsov, Osama Mohammed Afzal, Koen Dercksen, Nils Dycke, Alexander Goldberg, Tom Hope, Dirk Hovy, Jonathan K. Kummerfeld, Anne Lauscher, Kevin Leyton-Brown, Sheng Lu, Mausam, Margot Mieskes, Aurélie Névéol, Danish Pruthi, Lizhen Qu, Roy Schwartz, Noah A. Smith, Thamar Solorio, Jingyan Wang, Xiaodan Zhu, Anna Rogers, Nihar B. Shah, Iryna Gurevych. What Can Natural Language Processing Do for Peer Review?”, CoRR, Vol. abs/2405.06563, 2024.

Participants

- Osama Mohammed Afzal
MBZUAI – Abu Dhabi, AE
- Koen Dercksen
Radboud University
Nijmegen, NL
- Nils Dycke
TU Darmstadt, DE
- Alexander Goldberg
Carnegie Mellon University –
Pittsburgh, US
- Iryna Gurevych
TU Darmstadt, DE
- Jason Hartline
Northwestern University –
Evanston, US
- Tom Hope
The Hebrew University of
Jerusalem, IL
- Dirk Hovy
Bocconi University – Milan, IT
- Eddie Kohler
Harvard University – Allston, US
- Jonathan Kummerfeld
The University of Sydney, AU
- Ilia Kuznetsov
TU Darmstadt, DE
- Anne Lauscher
Universität Hamburg, DE
- Kevin Leyton-Brown
University of British Columbia –
Vancouver, CA
- Sheng Lu
TU Darmstadt, DE
- Dorsa Majdi
Sharif University of Technology –
Tehran, IR
- Mausam
Indian Institute of Technology –
New Delhi, IN
- Bahar Mehmani
Elsevier BV – Amsterdam, NL
- Margot Mieskes
Hochschule Darmstadt, DE
- Aurélie Névéol
CNRS – Orsay, FR
- Danish Pruthi
Indian Institute of Science –
Bangalore, IN
- Lizhen Qu
Monash University –
Clayton, AU
- Anna Rogers
IT University of
Copenhagen, DK
- Roy Schwartz
The Hebrew University of
Jerusalem, IL
- Nihar Shah
Carnegie Mellon University –
Pittsburgh, US
- Noah A. Smith
University of Washington –
Seattle, US
- Tamar Solorio
MBZUAI – Abu Dhabi, AE
- Jingyan Wang
Georgia Institute of Technology –
Atlanta, US
- Xiaodan Zhu
Queen's University –
Kingston, CA

