DAGSTUHL
REPORTS

**Volume 8, Issue 1, January 2018**

**ISSN 2192-5283**

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

*Publication date*
August, 2018

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



Digital Object Identifier: 10.4230/DagRep.8.1.i

*Aims and Scope*
The periodical *Dagstuhl Reports* documents the program and the results of Dagstuhl Seminars and Dagstuhl Perspectives Workshops.
In principal, for each Dagstuhl Seminar or Dagstuhl Perspectives Workshop a report is published that contains the following:

- an executive summary of the seminar program and the fundamental results,
- an overview of the talks given during the seminar (summarized as talk abstracts), and
- summaries from working groups (if applicable).

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



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

Report from Dagstuhl Seminar 18021

# Symmetric Cryptography

**Edited by**

**Joan Daemen[1], Tetsu Iwata[2], Nils Gregor Leander[3], and Kaisa Nyberg[4]**

1 **Radboud University Nijmegen, NL, and STMicroelectronics – Diegem, BE,** joan@cs.ru.nl
2 **Nagoya University, JP,** iwata@cse.nagoya-u.ac.jp
3 **Ruhr-Universität Bochum, DE,** gregor.leander@rub.de
4 **Aalto University, FI,** kaisa.nyberg@aalto.fi

## Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 18021 "Symmetric Cryptography", which was held on January 7–12, 2018 in Schloss Dagstuhl – Leibniz Center for Informatics. The seminar was the sixth in a series of Dagstuhl seminars on "Symmetric Cryptography", previously held in 2007, 2009, 2012, 2014, and 2016.

During the seminar, many of the participants presented their current research in the design, analysis, and application of symmetric cryptographic algorithms, including ongoing work and open problems. This report documents the abstracts or extended abstracts of the talks presented during the seminar, as well as summaries of the discussion sessions.

**Seminar** January 7–12, 2018 – https://www.dagstuhl.de/18021
**2012 ACM Subject Classification** Security and privacy → Cryptography → Cryptanalysis and other attacks, Security and privacy → Cryptography → Symmetric cryptography and hash functions
**Keywords and phrases** symmetric cryptography, cryptanalysis, authenticated encryption, cryptography for IoT, mass surveillance
**Digital Object Identifier** 10.4230/DagRep.8.1.1
**Edited in cooperation with** Maria Eichlseder

## 1 Executive Summary

*Nils Gregor Leander (Ruhr-Universität Bochum, DE)*
*Joan Daemen (Radboud University Nijmegen, NL, and STMicroelectronics – Diegem, BE)*
*Tetsu Iwata (Nagoya University, JP)*
*Kaisa Nyberg (Aalto University, FI)*



IT Security plays an increasingly vital role in everyday life and business. When talking on a mobile phone, when withdrawing money from an ATM or when buying goods over the internet, security plays a crucial role in both protecting the user and in maintaining public confidence in the system. Especially after the disclosure of the NSA's world-spanning spying activities and in the context of the Internet of Things, IT Security and privacy protection is a vital topic of the 21st century. In the Internet of Things (IoT) era, everything will be connected. Intel estimates that 200 billion objects will be connected by 2020. The objects

include for instance smart devices for healthcare, industrial control systems, automotive, and smart homes. Virtually all modern security solutions rely on cryptography.

Symmetric cryptography deals with the case that both the sender and the receiver of a message are using the same key. This differentiates symmetric cryptography from its asymmetric counterpart, where senders or verifiers use a "public key" and receivers or signers use a corresponding but different "private key". As asymmetric primitives are typically orders of magnitude less efficient than symmetric cryptographic schemes, symmetric cryptosystems remain the main workhorses of cryptography and highly relevant not only for academia, but also for industrial research and applications. While great progress has been made in designing and analyzing ciphers, fundamental aspects of these ciphers are still not fully understood. Moreover, as we have learned from the Snowden revelations, cryptography in general and symmetric cryptography in particular faces new fascinating challenges.

### Current Topics and Challenges

We identified the following three areas as among the most important topics for future research.

**Cryptography for the IoT.** Motivated by the upcoming IoT, one of the strong research trends in symmetric cryptography is about lightweight cryptography. Here, lightweight cryptography refers to strong cryptography, that can be executed on heavily resource constrained devices. Those efforts resulted in a wide variety of block cipher designs suitable for IoT applications. For instance, PRESENT designed in 2007 is one of the early designs with strong implementation advantages on hardware, and there have been other innovative follow-up block cipher designs. Some of them are standardized as the international standard, and used in thousands of devices in our daily lives. However, a block cipher is not the solution to all cryptographic purposes. For instance, to encrypt a certain amount of data, the block cipher has to be integrated into a suitable mode of operation. In most practical use cases, confidentiality is not the only concern, as many scenarios require data authenticity as well. Here a message authentication code (MAC) can be used to ensure authenticity. Authenticated encryption (AE) is used for protecting both confidentiality and authenticity.

The first MAC, called Chaskey, that specifically targets applications for lightweight cryptography was proposed only recently in 2014. The CAESAR project, an international competition for AE initiated at Dagstuhl, attracted several submissions that were designed for the purposes for lightweight cryptography. There is also a recent attempt to design a lightweight tweakable block cipher, an advanced primitive of a block cipher that allows more flexible usage, which can be efficiently integrated into highly secure encryption and/or authentication mechanisms. However, this research just started and many primitives and modes of operations suitable for lightweight crypto remain to be explored.

**Statistical Attacks.** Statistical attacks have been deployed widely and providing strong resistance against them has resulted in several important design criteria for contemporary symmetric primitives. The first type of statistical attacks that is applicable to a large set of block ciphers is differential cryptanalysis, introduced by Biham and Shamir. Since its invention in the early nineties several variants, tweaks and generalizations have been proposed and applied to many block ciphers. The second generally applicable attack on block ciphers is Matsui's linear cryptanalysis. Similarly to differential attacks, since its introduction, many extensions and improvements have been made. One main issue that has become apparent only recently is the accuracy of the underlying statistical models that researchers are using. Typically, those models are presented under some simplifying assumptions, whose validity remains an open question. It is an important challenge to settle these unsatisfactory

simplifications. This becomes even more important when the attacks are hard or impossible to verify experimentally due to the large computational costs involved. Moreover, to allow comparison between different attacks the researchers must agree on common attack models and parameters that measure the performance of the attack.

**Symmetric Cryptography and Real-World Needs.** The symmetric cryptography community has many very talented people and the state of the area has moved from it infancy in the seventies to a mature field today. However, we should ensure that the world's population does benefit of this progress. In particular, the Snowden leaks have painfully illustrated that citizen privacy and anonymity is next to non-existent nowadays. Secret services and IT corporations massively spy on people's communication and data storage for motives such as profit and surveillance. They don't seem to be hindered significantly in this at all by the pervasive deployment of cryptography (TLS, GSM, WPA, etc.). Cynically, monopolistic corporations like Google use encryption to protect the data of their users from prying eyes of other players such as network providers. It appears that much of the cryptography deployed today is there to protect the powers that be rather than protect human rights. With the roll-out of smart grid and internet-of-things surveillance will become quasi universal with all imaginable devices reporting on our behavior to big corporations. This situation has been addressed in several invited talks by Bart Preneel and Adi Shamir and they rightfully say that we as a cryptographic community should attempt to improve this. Along the same lines, Phil Rogaway gave a highly acclaimed invited talk at Asiacrypt 2015 on the moral aspects on cryptographic research. He invites us to do some introspection and ask the question: are we doing the right thing?

We believe these questions are important also for the symmetric crypto community. While the problem is certainly not restricted to symmetric cryptography and probably cannot be solved by symmetric cryptography alone, we should consider it our moral duty to improve the situation.

## Seminar Program

The seminar program consists of presentations about the above topics, and relevant areas of symmetric cryptography, including new cryptanalytic techniques and new designs. Furthermore, there were discussion sessions. In "Discussion on CAESAR with focus on robustness", we discussed about the meaning and relevance of the term robustness in general and for the CAESAR competition in particular. In "Discussion on Mass Surveillance", a number of questions related to the real-world relevance of the symmetric crypto community and its research were discussed. For both discussions we provide summery of the questions and results.

## 2 Table of Contents

# 3 Overview of Talks

## 3.1 TMD tradeoffs on small-state stream ciphers

*Willi Meier (FH Nordwestschweiz – Windisch, CH)*



Design and analysis of stream ciphers whose state is smaller than double the key size (small-state stream ciphers) is not fully exploited yet. For small-state stream ciphers that continuously use the non-volatile key in the state update, a TMD-TO distinguisher is described. A new mode for stream ciphers that continuously involve the IV (instead of the key) is proposed. Arguments are provided that this mode can resist generic TMD-TOs.

## 3.2 Towards Low Energy Stream Ciphers

*Vasily Mikhalev (Universität Mannheim, DE)*


**Joint work of** Subhadeep Banik, Frederik Armknecht, Takanori Isobe, Willi Meier, Andrey Bogdanov, Yuhei Watanabe, Francesco Regazzoni

Energy optimization is an important design aspect of lightweight cryptography. Since low energy ciphers drain less battery, they are invaluable components of devices that operate on a tight energy budget such as handheld devices or RFID tags. At Asiacrypt 2015, Banik et. al. presented the block cipher family Midori which was designed to optimize the energy consumed per encryption and which reduces the energy consumption by more than 30 % compared to previous block ciphers. However, if one has to encrypt/decrypt longer streams of data, i.e. for bulk data encryption/decryption, it is expected that a stream cipher should perform even better than block ciphers in terms of energy required to encrypt.

In this work, we address the question of designing low energy ciphers. To this end, we first analyze for common stream cipher design components their impact on the energy consumption. Based on this, we give arguments why indeed stream ciphers allow for encrypting long data streams with less energy than block ciphers and validate our findings by implementations. Afterwards, we use the analysis results to identify energy minimizing design principles for stream ciphers.

## 3.3 An LFSR-based Proof of Work

*Frederik Armknecht (Universität Mannheim, DE)*


**Joint work of** Frederik Armknecht, Ludovic Barman, Jens-Matthias Bohli, Ghassan O. Karame
**Main reference** Frederik Armknecht, Ludovic Barman, Jens-Matthias Bohli, Ghassan O. Karame: "Mirror: Enabling Proofs of Data Replication and Retrievability in the Cloud", in Proc. of the 25th USENIX Security Symposium, USENIX Security 16, Austin, TX, USA, August 10-12, 2016., pp. 1051–1068, USENIX Association, 2016.
**URL** https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/armknecht

In this talk, we present a novel cryptographic mechanism that is based on LFSRs (linear feedback shift register). It may be seen as a kind of proof of work in the following sense. The task is to compute certain elements from a finite group that are determined by an LFSR. The novel aspect here is that to this end, a short LFSR with small coefficients (over some finite field larger than GF(2)) are used but these specifications are kept secret. Instead, one publishes a related LFSR that is longer and has large coefficients. The aim of this scheme is that a prover who knows only the public specifications has to invest a minimum amount of effort to generate the elements while a verifier can use the knowledge of the secret specifications for a much faster verification.

The scheme has been initially introduced by us at USENIX Security 2016 to realize a scheme that allows for remote verification whether data has been stored with a sufficient level of redundancy. We think however that the presented mechanism can be of independent interest and poses some novel challenges, e.g., how to prove a minimum effort of the prover. In this talk, we explain the mechanism into more detail and also tell security arguments why a prover seem to have a higher computational effort than the verifier.

## 3.4 Rasta: Designing a cipher with low ANDdepth and few ANDs per bit

*Christoph Dobraunig (TU Graz, AT)*


**Joint work of** Christoph Dobraunig, Maria Eichlseder, Lorenzo Grassi, Virginie Lallemand, Gregor Leander, Florian Mendel, Christian Rechberger

Various lines of work have recently progressed with the investigation of the design and analysis of symmetric cryptographic schemes that minimize multiplications in one way or another. This has already led to unusual designs and interesting cryptanalytic insights. Even when only considering the class of schemes whose circuit has a natural and simple description in GF(2), there are various metrics that are interesting and useful: The total number of AND gates, the number of AND gates per encrypted bit, or the depth of the AND gate part of the circuit (ANDdepth), among others.

In this talk, we present with Rasta a design strategy for symmetric encryption that has ANDdepth d and at the same time only needs d ANDs per encrypted bit. The main result is that even for very low values of d between 2 and 6 we can give strong evidence that attacks may not exist. This contributes to a better understanding of the limits of what concrete symmetric-key constructions can theoretically achieve with respect to AND-related metrics.

## 3.5 Leakage-Resilient Authenticated Encryption

*Stefan Lucks (Bauhaus-Universität Weimar, DE)*



Practical cryptography often suffers from vulnerabilities to side-channel attacks. Two major approaches to deal with this problem are physical and algorithmic countermeasures. Physical countermeasures, such as "masking", try to prevent the side-channel, or to narrow it down. Algorithmic countermeasures are about meaningful security against adversaries with access to a (limited) side-channel.

This talk is about algorithmic countermeasures, which have been initiated with high hopes (Micali, Reyzin, 2004; Dziembowski, Pietrzak, 2008), but so far failed to take off in practice, specifically in Symmetric Cryptography. This talk is about algorithmic countermeasures and schemes which are supposed to be practically useful, while still maintaining a sound theoretical security proof. The formal approach is the introduction of leaking queries for (otherwise) ideal block ciphers.

## 3.6 Key Prediction Security of Keyed Sponges

*Bart Mennink (Radboud University Nijmegen, NL)*



The keyed sponge is a well-accepted method for message authentication. It processes data at a certain rate by sequential evaluation of an underlying permutation. If the key size $k$ is smaller than the rate, currently known bounds are tight, but if it exceeds the rate, state of the art only dictates security up to $2^{k/2}$. We take closer inspection at the key prediction security of the sponge and close the remaining gap in the existing security analysis: we confirm key security up to close to $2^k$, regardless of the rate. The result impacts all applications of the keyed sponge and duplex that process at a rate smaller than the key size, including the STROBE protocol framework, as well as the related constructions such as HMAC-SHA-3 and the sandwich sponge.

## 3.7 Tree-searching for trail bounds

*Gilles Van Assche (STMicroelectronics – Diegem, BE)*


**Joint work of** Silvia Mella, Joan Daemen, Gilles Van Assche
**Main reference** Silvia Mella, Joan Daemen, Gilles Van Assche: "New techniques for trail bounds and application to differential trails in Keccak", IACR Trans. Symmetric Cryptol., Vol. 2017(1), pp. 329–357, 2017.
**URL** http://dx.doi.org/10.13154/tosc.v2017.i1.329-357

In this presentation, we present the advantages of using a unit-based tree search for bounding the weight of differential and linear trails in cryptographic primitives. After recalling the definitions as set out in [1], we motivate the technique for the generation and the extension of differential and linear trails in the KECCAK-$f$ permutation. We then explain how the technique can easily avoid generating states that are equivalent under symmetry properties,

and how to use it to express trail extension when the states form an affine space. As an additional application, we show the bounds obtained on the new Xoodoo permutation and how the definition of units differed from those in Keccak-$f$. Finally, we conclude with questions that guide the application of the technique to a given cryptographic primitive.

### References

1 S. Mella, J. Daemen and G. Van Assche. New techniques for trail bounds and application to differential trails in Keccak. IACR Trans. Symmetric Cryptol. 2017(1): 329-357 (2017)

## 3.8 Merkle Tree is not Optimal

*Dmitry Khovratovich (University of Luxembourg, LU)*



No abstract given.

## 3.9 Fast Correlation Attack Revisited

*Yosuke Todo (NTT – Tokyo, JP)*


**Joint work of** Takanori Isobe, Willi Meier, Kazumaro Aoki, Bin Zhang

A fast correlation attack (FCA) is a well-known cryptanalysis technique for LFSR-based stream ciphers. The correlation between the initial state of an LFSR and corresponding key stream is exploited, and the goal is to recover the initial state of the LFSR. In this talk we revisit the FCA from a new point of view based on a finite field, and it brings a new property for the FCA when there are multiple linear approximations. Moreover we propose a novel algorithm by using the new property, which enables us to reduce both time and data complexities. We finally apply this technique to the Grain family, which is a well-analyzed class of stream ciphers. There are three stream ciphers, Grain-128a, Grain-128, and Grain-v1 in the Grain family, and Grain-v1 is in the eSTREAM portfolio and Grain-128a is standardized by ISO/IEC. As a result we break them all, and especially for Grain-128a, the cryptanalysis on its full version is reported for the first time.

## 3.10 Towards Quantitative Analysis of Cyber Security

*Adi Shamir (Weizmann Institute – Rehovot, IL)*


**Joint work of** Achiya Bar-On, Itai Dinur, Orr Dunkelman, Rani Hod, Nathan Keller, Eyal Ronen, Adi Shamir

Cyber security is a hot research area, but almost all the discussion about it is qualitative rather than quantitative. In this talk we consider the specific subtopic of backup schemes designed to protect computer systems against ransomware and cyber attacks. We develop a precise model with a concrete cost function, which describes the problem as an online/offline

optimization problem whose solution can be described by a pebbling game. We provide optimal backup schemes for all the cases with up to 10 backup devices, and find matching upper and lower bounds on the asymptotic efficiency of optimal backup schemes with an arbitrarily large number of backup devices.

## 3.11 Security of Caesar Candidates against (beyond) Birthday and/or Nonce-Reusing Attacks

*Damian Vizár (EPFL – Lausanne, CH)*


**Joint work of** Serge Vaudenay, Damian Vizár
**Main reference** Serge Vaudenay, Damian Vizár: "Under Pressure: Security of Caesar Candidates beyond their Guarantees", IACR Cryptology ePrint Archive, Vol. 2017, p. 1147, 2017.
**URL** http://eprint.iacr.org/2017/1147

The Competition for Authenticated Encryption: Security, Applicability and Robustness (CAESAR) has as its official goal to "identify a portfolio of authenticated ciphers that offer advantages over [the Galois-Counter Mode with AES]" and are suitable for widespread adoption." Each of the 15 candidate schemes competing in the currently ongoing 3$^{\text{rd}}$ round of CAESAR must clearly declare its security claims, i.a. whether or not it can tolerate nonce misuse, and what is the maximal data complexity for which security is guaranteed. These claims appear to be valid for all 15 candidates.

Interpreting "Robustness" in CAESAR as the ability to mitigate damage when security guarantees are void, we describe attacks with 64-bit complexity or beyond, and/or with nonce reuse for each of the 15 candidates. We then classify the candidates depending on how powerful does an attacker need to be to mount (semi-)universal forgeries, decryption attacks, or key recoveries. Rather than invalidating the security claims of any of the candidates, our results provide an additional criterion for evaluating the security that candidates deliver, which can be useful for e.g. breaking ties in the final CAESAR discussions.

## 3.12 Key-Recovery Attacks on Full Kravatte

*Henri Gilbert (ANSSI – Paris, FR)*


**Joint work of** Colin Chaigneau, Thomas Fuhr, Henri Gilbert, Jian Guo, Jérémy Jean, Jean-René Reinhard, Ling Song
**Main reference** Colin Chaigneau, Thomas Fuhr, Henri Gilbert, Jian Guo, Jérémy Jean, Jean-René Reinhard, Ling Song: "Key-Recovery Attacks on Full Kravatte", IACR Trans. Symmetric Cryptol., Vol. 2018(1), pp. 5–28, 2018.
**URL** http://dx.doi.org/10.13154/tosc.v2018.i1.5-28

We present a cryptanalysis of the July 2017 version of the full Kravatte and of a strengthened version presented in November at ECC 2017. Kravatte is an instantiation of the Farfalle construction of a pseudorandom function (PRF) with variable input and output length. This construction, proposed by Bertoni et al., represents an efficiently parallelizable and extremely versatile building block for the design of symmetric mechanisms, e.g. message authentication codes or stream ciphers. It relies on a set of non-linear permutations and on so-called rolling functions and can be split into a compression layer followed by a two-step expansion layer.

Kravatte instantiates Farfalle using linear rolling functions and non-linear permutations obtained by iterating the Keccak round function.

We develop several key recovery attacks against this PRF, based on three different attack strategies that bypass part of the construction and target a reduced number of permutation rounds. A higher order differential attack exploits the possibility to build an affine space of values in the cipher state after the compression layer. An algebraic meet-in-the-middle attack can be mounted on the second step of the expansion layer. Finally, a linear recurrence distinguisher can be found on intermediate states of the second step of the expansion layer and leveraged to mount a third attack. All the attacks rely on the ability to invert a small number of the final rounds of the construction. In particular, the last two rounds of the construction together with the final masking by the key can be algebraically inverted, which allows to recover the key. The complexities of the attacks are far below the claimed security level. Following the communication of the above cryptanalyses to the designers, a tweaked version of Kravatte was released in December 2017, in which one of the linear rolling functions is replaced by a non-linear rolling function.

## 3.13 Clustering Related-Tweak Characteristics

*Maria Eichlseder (TU Graz, AT)*


**Joint work of** Maria Eichlseder, Daniel Kales

The TWEAKEY/STK construction is an increasingly popular approach for designing tweakable block ciphers that notably uses a linear tweakey schedule. Several recent attacks have analyzed the implications of this approach for differential cryptanalysis and other attacks that can take advantage of related tweakeys. We generalize the clustering approach of a recent differential attack on the tweakable block cipher MANTIS-5 and describe a tool for efficiently finding and evaluating such clusters. More specifically, we consider the set of all differential characteristics compatible with a given truncated characteristic, tweak difference, and optional constraints for the differential. We refer to this set as a semi-truncated characteristic and estimate its probability by analyzing the distribution of compatible differences at each step.

We apply this approach to find a semi-truncated differential characteristic for MANTIS-6 with probability about $2^{-68}$ and derive a key-recovery attack with a complexity of about $2^{55}$ chosen-plaintext queries and computations. The data-time product is about $2^{110} \ll 2^{126}$.

## 3.14 Conditional Linear Cryptanalysis

*Stav Perle (Technion – Haifa, IL) and Eli Biham (Technion – Haifa, IL)*



In this talk we introduce an extension of linear cryptanalysis that may reduce the complexity of attacks by conditioning linear approximations on other linear approximations. We show that the bias of some linear approximations may increase under such conditions, so that after discarding the known plaintexts that do not satisfy the conditions, the bias of the remaining

known plaintexts increases. We show that this extension can lead to improvements of attacks, which may require fewer known plaintexts in total. By a careful application of our extension to Matsui's attack on the full 16-round DES we succeed to reduce the complexity of the best attack on DES to less than $2^{42}$.

## 3.15 Linear Cryptanalysis Using Low-Bias Approximations

*Tomer Ashur (KU Leuven, BE)*


**Joint work of** Tomer Ashur, Daniël Bodden, Orr Dunkelman
**Main reference** Tomer Ashur, Daniël Bodden, Orr Dunkelman: "Linear Cryptanalysis Using Low-bias Linear Approximations", IACR Cryptology ePrint Archive, Vol. 2017, p. 204, 2017.
**URL** http://eprint.iacr.org/2017/204

This work deals with linear approximations having absolute bias smaller than $2^{-n/2}$ which were previously believed to be unusable for a linear attack. We show how a series of observations which are individually not statistically significant can be used to create a $\chi^2$ distinguisher. This is different from previous works which combined a series of significant observations to reduce the data complexity of a linear attack.

## 3.16 Multidimensional, Affine and Conditional Linear Cryptanalysis

*Kaisa Nyberg (Aalto University, FI)*



Recently, new variants of linear cryptanalysis have been proposed. In this talk we focus on the affine multidimensional cryptanalysis and the conditional linear cryptanalysis. The affine method is based on multidimensional linear cryptanalysis and offers the option of discarding a whole half-space of linear approximations that do not contribute to statistical nonrandomness to keep only the information extracted from an affine subspace of linear approximatons. The conditional linear cryptanalysis was invented by Biham and Perle. In this talk we compare these methods and explain there relationships in the light of a small practical example originating from the DES cipher.

### Introduction

Linear cryptanalysis is a statistical method used for distinguishing a block cipher from a random family of permutations and can be extended to key recovery attacks in practical ciphers. It makes use of nonrandom behavior of linear approximations, which are single-bit values obtained by exclusive-or summation of certain input bits and output bits of the block cipher, or some rounds of it, over a large number of plaintexts.

Correlations of linear approximations over a block cipher with a fixed key are typically not statistically independent when taken as random variables over the data space. Methods that explicitly measure such dependencies, and use them in statistical analysis, have been presented previously by Murphy in [5] and very recently by Biham and Perle [1]. On the other hand, the main motivation of multidimensional linear cryptanalysis is that the

dependencies of linear approximations need not be measured explicitly as they are captured by the multidimensional linear test statistic. In this paper, we will present a concrete example to illustrate how this works in practice.

Next we briefly recall the multidimensional linear method, the affine space method, and the conditional linear cryptanalysis, and illustrate them for an example presented by Biham and Perle.

### Multidimensional Linear Cryptanalysis

In the context of linear cryptanalysis, a linear approximation of a transformation $F$ from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$ is a Boolean function in $\mathbb{F}_2^n$ defined by two vectors $a, \in \mathbb{F}_2^n$ and $b \in \mathbb{F}_2^n$ as follows

$$x \mapsto a \cdot x + b \cdot F(x).$$

In the statistical setting, a linear approximation is considered as a binary random variable $X$ over the given space of transformations with a probability density function defined by

$$\Pr(X = 0) = 2^{-n} \#\{x \in \mathbb{F}_2^n \,|\, a \cdot x + b \cdot F(x) = 0\}.$$

So we can write $X = a \cdot x + b \cdot F(x)$. In the linear-algebraic setting, a linear approximation $a \cdot x + b \cdot F(x)$ is identified with the vector $(a, b)$, called a mask pair, in the linear space $\mathbb{F}_2^n \times \mathbb{F}_2^m$ over $\mathbb{F}_2$.

Each linear approximation of $F(x)$ is a Boolean function and induces a probability distribution on $\{0, 1\}$. Its bias $\varepsilon_{(a,b)}$ is given by

$$\varepsilon_{(a,b)} = \Pr(a \cdot x + b \cdot F(x) = 0) - 1/2$$

and its correlation $c_{(a,b)}$ by

$$c_{(a,b)} = 2\varepsilon_{(a,b)} = \Pr(a \cdot x + b \cdot F(x) = 0) - \Pr(a \cdot x + b \cdot F(x) = 1).$$

Multidimensional linear cryptanalysis considers a number of linear approximations that form a linear subspace $V$ in $\mathbb{F}_2^n \times \mathbb{F}_2^m$. Let $t$ be the dimension of this subspace. Then a multidimensional linear approximation is a vector-valued Boolean function from $\mathbb{F}_2^n$ to $\mathbb{F}_2^t$. The components of this vector-valued function are in one-to-one correspondence with the mask pairs $(a, b) \in V$.

The strength of a multidimensional linear approximation is measured by its capacity $C_V$ given as follows

$$C_V = \sum_{(a,b) \in V, (a,b) \neq 0} c_{(a,b)}^2.$$

The multidimensional distinguisher is defined by the following test statistic

$$\begin{aligned} T(D) &= N \sum_{(a,b) \in V, (a,b) \neq 0} \hat{c}_{(a,b)}(D)^2, \text{ where} \\ D &= \text{ is a sample of } N \text{ plaintexts } x, \\ \hat{c}_{(a,b)}(D) &= N^{-1} \left( \#\{ x \in D \,|\, a \cdot x + b \cdot F(x) = 0\} - \#\{ x \in D \,|\, a \cdot x + b \cdot F(x) = 1\} \right). \end{aligned}$$

Under the assumption that the data for the observed correlations are computed from $N$ independently and randomly drawn $x$, the test statistic $T(D)$ is a Pearson's chi square test statistic with $2^t - 1$ degrees of freedom. For large $N$ and for uniformly distributed data, $T(D)$ follows a central chi square distribution. In the case, where the sample is drawn from a

nonuniform distribution, it was argued in [4] based on [3] that $T(D)$ follows a noncentral chi square distribution with noncentrality parameter $NC_V$, where $C_V$ is the nonzero capacity of the multidimensional linear approximation applied to cipher.

Let us now present an example of a typical situation where the subspace $V$ contains many useless linear approximations. Suppose that a multidimensional linear approximation of a cipher is built around a set of mask pairs $(a, b)$, where $a$ is a fixed nonzero mask on the plaintext and the ciphertext masks $b$ vary within a linear subspace $B$. The least linear subspace to contain all such masks is $\{0, a\} \times B$. Then the correlations of the linear masks of the form $(0, b)$, $b \in B$ have correlation zero, and do not add to the capacity of the multidimensional linear approximation, but just make the linear approximation space larger. Clearly,

$$\{a, 0\} \times B = (\{a\} \times B) \cup (\{0\} \times B)\,.$$

The affine subspace method to be presented next allows to discard the useless linear approximations in $\{0\} \times B$ and exploit the useful ones in the affine subspace $\{a\} \times B$.

### Affine Multidimensional Linear Cryptanalysis

Given a multidimensional linear approximation as described in the previous section, we split $V$ into two halves, a subspace $U$ of dimension $s = t - 1$ and the affine subspace $V \setminus U$. Given $(a_1, b_1) \in V \setminus U$, all the mask pairs in $V$ can be written in the form $(a_2, b_2)$ or $(a_1 + a_2, b_1 + b_2)$, where $(a_2, b_2) \in U$.

First, Let us apply the multidimensional linear model to the sapce $V$. Then the test statistic $T_V(D)$ is computed as follows

$$T_V(D) \quad = \quad N \sum_{(a,b)\in V} \hat{c}_{(a,b)}(D)^2.$$

Secondly, let us apply the multidimensional model to the linear approximations in the subspace $U$ to obtain whence the test statistic is computed as

$$T_U(D) \quad = \quad N \sum_{(a_2,b_2)\in U} \hat{c}_{(a_2,b_2)}(D)^2.$$

We now define the affine test statistic $T_{\text{aff}}(D)$ as follows

$$T_{\text{aff}}(D) = T_V(D) - T_U(D) = N \sum_{(a_2,b_2)\in U} \hat{c}_{(a_1+a_2,b_1+b_2)}(D)^2.$$

Under the assumption that the data for the observed correlations are computed from $N$ independently and randomly drawn $x$, we obtain using Pearson's chi square test that $T_{\text{aff}}(D)$ is chi square distributed with $2^s$ degrees of freedom, for large $N$. In the random case, we then have a central chi square distribution with mean $2^s$ and variance $2^{s+1}$. Otherwise, the mean can be computed from the expression $T_{\text{aff}}(D) = T_V(D) - T_U(D)$ to get

$$\text{Exp}\, T_{\text{aff}}(D) \quad = \quad 2^s + N(C_V - C_U).$$

Thus the noncentrality parameter of the chi square distribution of $T_{\text{aff}}(D)$ in the cipher case is equal to $C_V - C_U$, and we obtain

$$\text{Var}\, T_{\text{aff}}(D) \quad = \quad 2\left(2^s + 2N(C_V - C_U)\right).$$

Similarly as for multidimensional linear cryptanalysis, the derived affine statistical model can be used in cryptanalytic distinguishing and key-recovery attacks. Next we present a second example which shows that the affine space method can improve upon the multidimensional linear cryptanalysis.

### Example of Biham and Perle

Recently, Eli Biham and Stav Perle proposed a new cryptanalysis method called as conditional linear cryptanalysis [1]. It applies to the case where two linear approximations are mutually dependent. For example, they found two dependent linear approximations in DES. We denote the random variables related to them by $X$ and $Y$. They have the following probability density functions

$$\Pr(X=0)=\frac{1}{2}+\varepsilon \qquad \Pr(X=1)=\frac{1}{2}-\varepsilon$$
$$\Pr(Y=0)=\frac{1}{2} \qquad \Pr(Y=1)=\frac{1}{2}.$$

Their dependency is given in terms of conditional probabilities

$$\Pr(X=0|Y=0)=\frac{1}{2}+2\varepsilon, \qquad \Pr(X=0|Y=1)=\frac{1}{2},$$
$$\Pr(X=1|Y=0)=\frac{1}{2}-2\varepsilon, \qquad \Pr(X=1|Y=1)=\frac{1}{2}.$$

We use this example to illustrate the behavior of the three variants of linear cryptanalysis.

**The multidimensional linear model.** The capacity of the 2-dimensional multidimensional linear approximation in $V$ spanned by the linear approximations $X$ and $Y$ is equal to

$$C_V \quad = \quad c_X^2 + c_Y^2 + c_{X+Y}^2.$$

Note that we use the variable symbol instead of the mask pairs to identify the non-zero linear approximations. It is easy to check that the linear approximation $X+Y$ has the same bias as $X$, and the bias of $Y$ is equal to zero. We get $C_V = 8\varepsilon^2$. Then the multidimensional test statistic

$$T_V = N\left(\hat{c}_X(D)^2 + \hat{c}_Y(D)^2 + \hat{c}_{X+Y}(D)^2\right)$$

has a noncentral chi square distribution with 3 degrees of freedom and noncentrality parameter equal to $8N\varepsilon^2$.

**The affine linear model.** Since $c_Y = 0$, it does not contribute to the capacity of the multidimensional distribution. To discard it, we apply the affine linear model with the 1-dimensional subspace $U = \{0, Y\}$. Then the affine test statistic

$$T_{\text{aff}} = N\left(\hat{c}_X(D)^2 + \hat{c}_{X+Y}(D)^2\right)$$

has chi square distribution with 2 degrees of freedom and noncentrality parameter

$$N(C_V - C_U) = N\left(c_X^2 + c_{X+Y}^2\right) = 8N\varepsilon^2.$$

It means that the affine linear test has the same noncentrality parameter but less degrees of freedom than the multidimensional linear test and hence is more efficient.

**The conditional linear model.** Recently, Biham and Perle proposed conditional linear cryptanalysis to exploit high conditional correlations [1]. The idea is to use the analogical statistical model as for classical linear cryptanalysis in the context of conditional probabilities and biases by discarding the data that does not satisfy the condition. According to this model the observed number of data $\hat{N}'$ that satisfy $X=0$ within a sample of $N'$ plaintext-ciphertext pairs that satisfy $Y=0$ is binomially distributed with probability $\Pr(X=0\,|\,Y=0) =$

$1/2+2\varepsilon$ and sample size $N'$. The bias of this conditional distribution is $2\varepsilon$ and the correlation is $4\varepsilon$. Hence the distribution of the observed correlation

$$2\hat{N}'/N' - 1$$

can be approximated by a normal distribution with mean $c_{X|_{Y=0}} = 4\varepsilon$ and variance $1/N'$, where we denoted by

$$c_{X|_{Y=0}} = \Pr(X = 0 \,|\, Y = 0) - \Pr(X = 1 \,|\, Y = 0)$$

the conditional correlation.

The data complexity estimate obtained from the normal distribution is the same that can be obtain using the chi square distribution obtained from the squared observed correlation [2]. More precisely, the conditional test statistic $T_{\text{cond}}$ defined as

$$T_{\text{cond}} = N'(2\hat{N}'/N' - 1)^2 \sim \chi_1^2(\delta)$$

where

$$\delta = N' c_{X|Y=0}^2 = 16N'\varepsilon^2$$

gives the same data complexity estimate as the binomial (normal) test statistic $\hat{N}'/N'$ traditionally used in linear cryptanalysis. Since $Y$ is unbiased, it is estimated that for the total size $N$ of the sample is equal to $2N'$.

We can see that the non-centrality parameter $\delta$ is the same also in the case of conditional linear cryptanalysis. To explain this coincidence, we need to express the capacity of the affine linear approximation in terms of the probabilities $p_{00}$, $p_{01}$, $p_{10}$, and $p_{11}$, where

$$p_{uv} = \Pr(X = u, Y = v),\ u = 0, 1 \text{ and } v = 0, 1$$

are the probabilities of the 2-dimensional variable $(X, Y)$. Then it can be shown that

$$C_V - C_U = c_X^2 + c_{X+Y}^2 = 2\left((p_{00} - p_{10})^2 + (p_{01} - p_{11})^2\right)$$

Now we observe that $p_{01} - p_{11} = 0$. It means that all the nonbalancedness of the distribution of this pair $(X, Y)$ of linear approximations can measured by the first term

$$p_{00} - p_{10} = \Pr(Y = 0)(\Pr(X = 0 \,|\, Y = 0) - \Pr(X = 1 \,|\, Y = 0)) = \Pr(Y = 0) c_{X|_{Y=0}},$$

that is, by the product of $\Pr(Y = 0)$ and the conditional correlation $c_{X|_{Y=0}}$.

Finally, we observe that the conditional approach allows to reduce the degree of freedom to one while keeping the noncentrality parameter the same as in the usual multidimensional cryptanalysis and in the affine multidimensional cryptanalysis. We conclude that from the three statistical models considered for the given example, the conditional linear cryptanalysis of Biham and Perle gives the most efficent statistical distinguisher.

**Acknowledgements.** I wish to thank Eli Biham for discussions related to conditional linear cryptanalysis and Céline Blondeau for suggestions how to improve the presentation.

## References

1 Eli Biham and Stav Perle. Conditional linear cryptanalysis. Presentation at Romanian Cryptology Days, Bucharest, Romania, Sept 18–20, 2017, and at Dagstuhl Seminar 18021 “Symmetric Cryptography”, Jan 7–12, 2018.

2 Céline Blondeau and Kaisa Nyberg. Improved parameter estimates for correlation and capacity deviates in linear cryptanalysis. IACR Transactions on Symmetric Cryptology, 2016(2):162–191, 2017.
3 F.C. Drost, W.C.M. Kallenberg, D.S.Moore, and J.Oosterhoff. Power Approximations to Multinomial Tests of Fit. Journal of the American Statistican Association, 84(405):130–141, Mar 1989.
4 Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional Extension of Matsui's Algorithm 2. In Orr Dunkelman, editor, FSE 2009, vol. 5665 of LNCS, pages 209–227. Springer, 2009.
5 Sean Murphy. The independence of linear approximations in symmetric cryptanalysis. IEEE Transactions on Information Theory, 52(12):5510–5518, 2006.

## 3.17 The Chi-Squared Method

*Stefano Tessaro (University of California – Santa Barbara, US)*


**Joint work of** Wei Dai, Viet Tung Hoang

Proving tight bounds on information-theoretic indistinguishability is a central problem in symmetric cryptography. In this talk, I introduce a new method for information-theoretic indistinguishability proofs, called "the chi-squared method". At its core, the method requires upper-bounds on the so-called chi-squared divergence between the output distributions of two systems being queried

I will showcase the chi-squared method by giving a simple proof of optimal security for the XOR of two random permutations, which improves upon bounds previously shown with much more involved machinery (e.g., mirror theory).

## 3.18 Some applications of the chi square method

*Mridul Nandi (Indian Statistical Institute – Kolkata, IN)*



In this talk, I would like to discuss some possible applications of chi-squared method. So far, it has been applied to the sum of random permutations, EDM and truncation of random permutation. Very recently, it is also applied to prove the PRF security of sum of permutation where the inputs are reused in a certain way. This is related the well known powerful tool – mirror theory. As the proof of the Mirror theory is highly complex and contains several non-trivial gap, it would be nice to explore other way out for the application of the mirror theory. Chi-squared method could be such an alternative. I also describe how to prove a weaker form of mirror theory using the chi-squared method result applied to the reused sum of permutation. Using this, I would be able to prove the weak-PRF full n bit security of EDM. This can be possibly extended to standard PRF security, but requires more closer analysis.

## 3.19 Beyond-Birthday-Bound Secure MACs

*Yannick Seurin (ANSSI – Paris, FR)*


**Joint work of** Benoît Cogliati, Tetsu Iwata, Jooyoung Lee, Kazuhiko Minematsu, Thomas Peyrin

A Message Authentication Code (MAC) is a fundamental symmetric primitive allowing two entities sharing a secret key to verify that a received message originates from one of the two parties and was not modified by an attacker. Most existing MACs are built from a block cipher, e.g., CBC-MAC or OMAC, or from a cryptographic hash function, e.g., HMAC. In general, MACs which are constructed from a block cipher are secure only up to the so-called birthday bound with respect to the block size $n$ of the block cipher: they become insecure when $\sim 2^{n/2}$ (blocks of) messages have been treated. This might be problematic, especially when relying on lightweight block ciphers with small block size or when updating the secret key is impractical. In this talk, we survey recent results on MAC constructions based on a block cipher or a tweakable block cipher which are secure beyond the birthday bound such as EWCDM [1], ZMAC [2] and HaT/NaT/HaK/NaK [3] and we highlight some open problems along the way.

### References

1 Benoît Cogliati and Yannick Seurin. EWCDM: An Efficient, Beyond-Birthday Secure, Nonce-Misuse Resistant MAC. In Matthew Robshaw and Jonathan Katz, editors, CRYPTO 2016 (1), vol. 9814 of LNCS, pp. 121–149. Springer, 2016.
2 Tetsu Iwata, Kazuhiko Minematsu, Thomas Peyrin, and Yannick Seurin. ZMAC: A Fast Tweakable Block Cipher Mode for Highly Secure Message Authentication. In Jonathan Katz and Hovav Shacham, editors, CRYPTO 2017 (3), vol. 10403 of LNCS, pp. 34–65. Springer, 2017.
3 Benoît Cogliati, Jooyoung Lee, and Yannick Seurin. New Constructions of MACs from (Tweakable) Block Ciphers. IACR Trans. Symmetric Cryptol., 2017(2):27–58, 2017.

## 3.20 Recent Advancements in Sponge-Based MACs

*Kan Yasuda (NTT - Tokyo, JP)*



No abstract given.

## 3.21 The collision-resistance of keyed hashing

*Joan Daemen (Radboud University Nijmegen, NL, and STMicroelectronics – Diegem, BE)*



MAC functions and pseudorandom functions with arbitrary input length often consist of two stages: a keyed hash function that compresses the input to a fixed-length *accumulator* followed by a function that maps the accumulator to the output, that may also have variable

length. The security requirement for the keyed hash is that it should be difficult for an adversary that does not know the key to find inputs that collide in the accumulator. More precisely, the adversary gets adaptive query access to the keyed hash function where she gets the image of the accumulator through a random oracle. In other words, she can only see whether inputs collide or not in the accumulator.

Keyed hash functions are implemented in a wide variety of ways: serial constructions such as CBC-MAC, polynomial evaluations in a finite field and Pelican-MAC or parallel constructions such as PMAC or Farfalle. These constructions make use of block ciphers, tweakable block ciphers or permutations. Each of these have their own advantages and disadvantages, but all are vulnerable to a generic collision attack that has success probability $M^2 2^{-(b+1)}$ with $b$ the size of the accumulator and $M$ the number of queries to the keyed hash function (data complexity).

One usually characterizes the level of security that a cryptographic scheme offers by the so-called *security strength* that is expressed in bits. For a certain attack, it is the binary logarithm of its data complexity $M$ minus that of its success probability $p$, so $s = \log_2 M - \log_2 p$. For the generic attacks, at one end of the spectrum is an attack with just a couple of queries that has $s \approx b$. At the other end the success probability approaches 1 when $M \approx 2^{b/2}$ and hence it has $s \approx b/2$. So the maximum achievable security strength decreases from $b$ to $b/2$ bits as the attack complexity grows from 2 to $2^{b/2}$. This curve is called the *birthday bound*.

When designing a keyed hash function, different strategies may be followed. First, one may aim either for a capacity claim or for a security strength claim.

In the former, one makes a claim for the function that there are no attacks with success probability below $M^2 2^{-(c+1)}$ with $c$ some specified constant usually called the *capacity*. In the so-called *hermetic* design strategy, one chooses $b = c$, implying that there are no attacks better than the generic attack and hence that the used primitive has no exploitable weaknesses. This usually requires using a primitive with a significant computational cost. This cost can be reduced drastically by taking $b > c$, so by over-dimensioning the primitive. An example of this strategy is Pelican-MAC, that has $c = 120$ and uses 4 unkeyed AES rounds as permutation, so $b = 128$.

In a security strength claim one states that there are no attacks with success probability below $M2^{-s}$, possibly putting an upper bound on $M$. If this upper bound is $2^a$ with $a < s$, this requires taking $b$ at least $s + a$ and $2s$ otherwise. An example is Kravatte with $b = 1600$ and $s = 137$.

Determining the best attack strategy for collision attacks for the different constructions in combination with different primitives is an interesting research problem and allows gaining insight in how to build the most efficient keyed hash function for some given set of target platforms and for some target security, either expressed by a capacity $c$ or a strength $s$.

## 3.22 Challenges and Opportunities for the Standardization of Threshold Cryptography

*Nicky Mouha (NIST – Gaithersburg, US)*


**Joint work of** Apostol Vassilev, Nicky Mouha, Luís Brandão
**Main reference** Apostol Vassilev, Nicky Mouha, Luis Brandao: "Psst, Can You Keep a Secret?", IEEE Computer, Vol. 51(1), pp. 94–97, 2018.
**URL** http://dx.doi.org/10.1109/MC.2018.1151029

Cryptography lies at the heart of the protection of data at rest and in transit over the Internet. The security of data afforded by the employed cryptographic primitives depends not only on their theoretical properties but also on the robustness of their implementations in software and hardware. Threshold cryptography introduces a computational paradigm that enables a higher level of assurance for the implementations of cryptographic primitives.

We discuss challenges and opportunities related to the standardization of threshold cryptography [1], and give some insights into their application to symmetric-key cryptography.

### References

1 Apostol Vassilev, Nicky Mouha, Luís Brandão. Psst, Can you Keep a Secret? IEEE Computer 51(1): 94–97, 2018, https://dx.doi.org/10.1109/MC.2018.1151029

## 3.23 Tools on Cryptanalysis

*Stefan Kölbl (Technical University of Denmark – Lyngby, DK)*


**Joint work of** Zahra Eskandari, Andreas Brasen Kidmose, Stefan Kölbl, Tyge Tiessen

The division property method is a powerful technique to determine integral distinguishers on block ciphers. While the complexity of finding these distinguishers is higher, it has recently been shown that MILP and SAT solvers can efficiently find such distinguishers.

In this work, we provide a framework to fully automate finding those distinguishers which solely relies on a simple description of the cryptographic primitive. We demonstrate the ease of use by finding integral distinguishers for more than 30 primitives based on different design strategies and present several new or improved distinguishers for ChaCha, ChasKey, DES, GIFT, LBlock, Mantis, Qarma, RoadRunner, Salsa and SM4.

## 3.24 A survey of recent results on AES permutations

*Christian Rechberger (TU Graz, AT)*



We survey recent results on new properties of AES, and subspace trail cryptanalysis as a way to describe it. This includes various properties of 5-round AES that hold for any secret key, and a 10-round property that holds for a set of $2^{32}$ chosen keys.

## 3.25 Cryptanalysis of Reduced Round AES, Revisited

*Orr Dunkelman (University of Haifa, IL)*


**Joint work of** Achiya Bar-On, Nathan Keller, Eyal Ronen, Adi Shamir

Determining the security of AES is a central problem in cryptanalysis, but progress in this area had been slow and only a handful of cryptanalytic techniques led to significant advancements. At Eurocrypt 2017 Grassi et al. presented a novel type of distinguisher for AES-like structures, but so far all the published attacks which were based on this distinguisher were either inferior or comparable to previously known attacks in their complexity. In this paper we combine the technique of Grassi et al. with several other techniques to obtain the best known key recovery attack on 5-round AES in the single-key model, reducing its data, memory and time complexities from about $2^{32}$ to about $2^{22.5}$. Extending our techniques to 7-round AES, we obtain the best known attacks which use practical amounts of data and memory, breaking the record for such attacks which was obtained 18 years ago by the classical Square attack.

## 3.26 Integral Attacks on AES

*Meiqin Wang (Shandong University – Jinan, CN)*



Reduced-round version of AES has been a popular underlying primitives to design new cryptographic schemes. The security including the distinguishing property of AES deserves to study more. Recently, the key-dependent integral and impossible differential distinguishers for 5-round AES have been put forward. Later, the structural distinguisher and Yoyo distinguisher for 5-round or 6-round AES have been introduced. Although the complexities of the key-dependent integral and impossible differential distinguishers are much higher than those of the structural or Yoyo distinguisher for 5-round AES, more detailed property for MixColumn can be identified by them. Traditional impossible differential and integral distinguishers for 4-round AES have approximately equal data complexity. However, for the recent proposed key-dependent distinguishers, there is a big gap between the complexities of the integral and impossible differential distinguishers. Even with the same property of MixColumn, the integral distinguisher requires the whole codebook while the impossible distinguisher just needs $2^{98.2}$ chosen plaintexts. Moreover, the complexities of traditional impossible differential or integral distinguishers are identical for the chosen-plaintext and chosen-ciphertext settings, but they are very different for the key-dependent distinguishers. Till now, the 5-round integral and impossible differential distinguishers can only work for chosen-ciphertext and chosen plaintext settings, respectively.

In this talk, by appending the condition for the output values for 5-round zero-correlation linear hull, we can transform such zero-correlation linear hull to a new key-dependent integral distinguisher for 5-round AES with $2^{96}$ chosen plaintexts which is much better than the previous integral distinguisher at CRYPTO 2015 with the whole codebook. Secondly, we focus on transforming the key-dependent impossible differential distinguishers from the chosen-plaintext to chosen-ciphertext situation by setting the condition on the output values.

We found that the key-dependent integral distinguishers have very different complexities but the key-dependent impossible differential distinguishers have no significant difference for the complexity under different attacking modes. Finally, we utilize our proposed 5-round integral distinguisher to recover the key for 6-round AES. Although the key recovery attack is no better than the previous attacks with 4-round distinguishers, it is the first integral key-recovery attack on 6-round based on 5-round distinguisher.

## 3.27 On Sboxes sharing the same DDT

*Anne Canteaut (INRIA – Paris, FR)*


**Joint work of** Christina Boura, Anne Canteaut, Jérémy Jean, Valentin Suder

In this work, we discuss two notions of differential equivalence on Sboxes. First, we introduce the notion of DDT-equivalence which applies to vectorial Boolean functions that share the same difference distribution table (DDT). It is worth noticing that this property equivalently means that the two functions share the same squared Walsh transform. Next, we compare this notion to what we call the $\gamma$-equivalence, applying to vectorial Boolean functions whose DDTs have the same support. This second property has been studied by Gorodilova for quadratic APN functions and in particular for the Gold family of functions. We discuss the relation between these two equivalence notions, demonstrate that the number of DDT- or $\gamma$-equivalent functions is invariant under EA- and CCZ-equivalence. This answers an open problem raised by Gorodilova. In parallel, we also provide an algorithm for computing the DDT-equivalence and the $\gamma$-equivalence classes of a given function. We study the sizes of these classes for some families of Sboxes.

## 3.28 Boomerang Connectivity Table (BCT) for Boomerang Attacks

*Yu Sasaki (NTT – Tokyo, JP)*


**Joint work of** Carlos Cid, Tao Huang, Thomas Peyrin, Ling Song

A boomerang attack is a cryptanalysis framework that regards a block cipher $E$ as the composition of two sub-ciphers $E_1 \circ E_0$ and builds a particular characteristic for $E$ with probability $p^2q^2$ by combining differential characteristics for $E_0$ and $E_1$ with probability $p$ and $q$, respectively. Crucially the validity of this figure is under the assumption that the characteristics for $E_0$ and $E_1$ can be chosen independently. Indeed, Murphy has shown that independently chosen characteristics may turn out to be incompatible. On the other hand, several researchers observed that the probability can be improved to $p$ or $q$ around the boundary between $E_0$ and $E_1$ by considering a positive dependency of the two characteristics, e.g. the ladder switch and S-box switch by Biryukov and Khovratovich. This phenomenon was later formalised by Dunkelman et al. as a sandwich attack that regards $E$ as $E_1 \circ E_m \circ E_0$, where $E_m$ satisfies some differential propagation among four texts with probability $r$, and the entire probability is $p^2q^2r$. In this paper, we revisit the issue of dependency of two characteristics in $E_m$, and propose a new tool called *Boomerang Connectivity Table (BCT)*,

which evaluates $r$ in a systematic and easy-to-understand way when $E_m$ is composed of a single S-box layer. With the BCT, previous observations on the S-box including the incompatibility, the ladder switch and the S-box switch are represented in a unified manner. Moreover, the BCT can detect a new switching effect, which shows that the probability around the boundary may be even higher than $p$ or $q$.

## 3.29 QCCA on Feistel

*Tetsu Iwata (Nagoya University, JP)*


**Joint work of** Gembu Ito, Tetsu Iwata, Ryutaroh Matsumoto

Kuwakado and Morii considered quantum chosen plaintext attacks and showed an efficient distinguishing attack against the three-round Feistel cipher by using Simon's period finding algorithm [1]. In this talk, we consider quantum chosen ciphertext attacks, and present an efficient distinguishing attack against the four-round Feistel cipher.

### References

1 Hidenori Kuwakado and Masakatu Morii. Quantum distinguisher between the 3-round Feistel cipher and the random permutation. ISIT 2010, pp. 2682–2685, IEEE, 2010.

## 3.30 Some Feistel structures with low degree round functions

*Arnab Roy (University of Bristol, GB)*



We consider several generalized Feistel constructions with low-degree round function. In particular, we study cases of the form $x \to x^r$ for various $r$, with focus on the simplest case $r = 3$. Our analysis allows us to propose more efficient generalizations of the MiMC design (Asiacrypt'16). We evaluate the new designs in three application areas. Whereas MiMC was not competitive at all in a recently proposed new class of PQ-secure signature scheme, our new construction leads to about 30 times smaller signatures than MiMC. For MPC use cases, where MiMC seems to outperform all other competitors to start with, we observe substantial improvements in throughput by a factor of around 5 and simultaneously a 10-fold reduction of pre-processing effort, at the cost of a higher latency. Another use case where MiMC already outperforms other designs, in the area of SNARKs, only sees modest improvements.

## 3.31 Generalized Feistel Networks with Optimal Diffusion

*Léo Paul Perrin (INRIA – Paris, FR)*


**Joint work of** Léo Perrin, Angela Promitzer, Sebastian Ramacher, Christian Rechberger
**Main reference** Léo Perrin, Angela Promitzer, Sebastian Ramacher, Christian Rechberger: "Improvements to the Linear Layer of LowMC: A Faster Picnic", IACR Cryptology ePrint Archive, Vol. 2017, p. 1148, 2017.
**URL** http://eprint.iacr.org/2017/1148

Generalized Feistel networks are a common block cipher structure. In [2], Suzaki and Minematsu introduced an improved branch permutation which allowed a faster diffusion in generalized Feistel networks. While such structures usually need $b$ rounds to achieve full diffusion over $b$ branches, Suzaki and Minematsu's requires only about $2\log_2(b)$.

In this talk, we presented a different method for building generalized Feistel networks with fast diffusion. The round function is simple: it can be seen as a simple two-branched Feistel network where the Feistel function consists in an S-Box layer followed by a rotation of the corresponding words. The core idea consists in using different rotation amounts in each round. Indeed, if those are chosen carefully then we can prove a fast diffusion. For example, if the rotation sequence is $\{0, 1, 0, 2, 0, 4, 0, 8, 0, ...\}$, then diffusion is essentially as fast as in [2]. Furthermore, if the sequence is instead the Fibonacci sequence $\{0, 1, 1, 2, 3, 5, 8, ...\}$, then diffusion is even faster and reaches an optimal bound first identified by Suzaki and Minematsu. The latter construction was used in [1] to build linear layers with full diffusion allowing a constant time implementation with a speed comparable to a table-based one.

### References

1 Léo Perrin, Angela Promitzer, Sebastian Ramacher, and Christian Rechberger. Improvements to the Linear Layer of LowMC: A Faster Picnic. Cryptology ePrint Archive, Report 2017/1148, 2017.
2 Tomoyasu Suzaki and Kazuhiko Minematsu. Improving the Generalized Feistel. FSE 2010, pp 19–39. Springer, 2010.

## 3.32 An Improved Affine Equivalence Algorithm

*Itai Dinur (Ben Gurion University – Beer Sheva, IL)*


**Main reference** Itai Dinur: "An Improved Affine Equivalence Algorithm for Random Permutations", IACR Cryptology ePrint Archive, Vol. 2018, p. 115, 2018.
**URL** https://eprint.iacr.org/2018/115

In this work we study the affine equivalence problem, where given two functions $\vec{F}, \vec{G} : \{0,1\}^n \to \{0,1\}^n$, the goal is to determine whether there exist invertible affine transformations $A_1, A_2$ over $GF(2)^n$ such that $\vec{G} = A_2 \circ \vec{F} \circ A_1$. Algorithms for this problem have several well-known applications in the design and analysis of Sboxes, cryptanalysis of white-box ciphers and breaking a generalized Even-Mansour scheme.

We describe a new algorithm for the affine equivalence problem and focus on the variant where $\vec{F}, \vec{G}$ are permutations over $n$-bit words, as it has the widest applicability. The complexity of our algorithm is about $n^3 2^n$ bit operations with very high probability whenever $\vec{F}$ (or $\vec{G}$) is a random permutation. This improves upon the best known algorithms for this problem (published by Biryukov et al. at EUROCRYPT 2003), where the first algorithm has

time complexity of $n^3 2^{2n}$ and the second has time complexity of about $n^3 2^{3n/2}$ and roughly the same memory complexity.

Our algorithm is based on a new structure (called a *rank table*) which is used to analyze particular algebraic properties of a function that remain invariant under invertible affine transformations. Besides its standard application in our new algorithm, the rank table is of independent interest and we discuss several of its additional potential applications.

## 3.33 Invariant Attacks and (Non-)linear Approximations

*Christof Beierle*


**Joint work of** Christof Beierle, Anne Canteaut, Gregor Leander

This work discusses nonlinear approximations for block cipher cryptanalysis by embedding it into the better-understood framework of linear cryptanalysis.

In the first part we show that, in some cases, a deterministic nonlinear approximation (aka. nonlinear invariant attack) over a keyed instance of a cipher implies the existence of a (non-trivial) highly-biased linear approximation over the same instance. In the second part, we present a framework for studying non-deterministic nonlinear approximations. In particular, by transforming the cipher under consideration by conjugating each keyed instance with a fixed permutation, we are able to transfer many methods from linear cryptanalysis to the nonlinear case. Using this framework we in particular show that there exist ciphers for which some transformed versions are significantly weaker with respect to linear cryptanalysis than their original counterparts. This suggests that the basic security argument of counting the minimum number of active S-boxes may not be sufficient to avoid such kind of attacks.

## 3.34 Recent results on reduced versions of Ketje

*Maria Naya-Plasencia (INRIA – Paris, FR)*


**Joint work of** Thomas Fuhr, María Naya-Plasencia, Yann Rotella
**Main reference** Thomas Fuhr, María Naya-Plasencia, Yann Rotella: "State-Recovery Attacks on Modified Ketje Jr", IACR Trans. Symmetric Cryptol., Vol. 2018(1), pp. 29–56, 2018.
**URL** http://dx.doi.org/10.13154/tosc.v2018.i1.29-56

In this article we study the security of the authenticated encryption algorithm Ketje against divide-and-conquer attacks. Ketje is a third-round candidate in the ongoing CAESAR competition, which shares most of its design principles with the SHA-3 hash function. Several versions of Ketje have been submitted, with different sizes for its internal state. We describe several state-recovery attacks on the smaller variant, called Ketje Jr. We show that if one increases the amount of keystream output after each round from 16 bits to 40 bits, Ketje Jr becomes vulnerable to divide-and-conquer attacks with time complexities $2^{71.5}$ for the original version and $2^{82.3}$ for the current tweaked version, both with a key of 96 bits. We also propose a similar attack when considering rates of 32 bits for the non-tweaked version. Our findings do not threaten the security of Ketje, but should be taken as a warning against potential future modifications that would aim at increasing the performance of the algorithm.

## 3.35 On the security of LINE messaging application

*Kazuhiko Minematsu*


**Joint work of** Takanori Isobe, Kazuhiko Minematsu

In this talk, we study the security of LINE messaging application (a.k.a. text messaging or instant messaging). LINE is by far the common messaging application in Japan, and is also popular in some East Asian countries, such as Taiwan, Thailand and Indonesia. There are 217 million monthly active users, as of Jan. 2017.

LINE provides an End-to-End (E2E) encryption scheme called Letter Sealing since 2015. After the reverse engineering work on Letter Sealing by Curtiss [1], LINE corporation has published a whitepaper [3] describing the specification of Letter Sealing in 2016. Recently, Espinoza et. al [2] proposed a replay attack against Letter Sealing.

We investigated this whitepaper, and found several vulnerabilities not covered by prior work. With these vulnerabilities, we found practical attacks against LINE's one-to-one messaging and group messaging. The vulnerabilities are listed as follows.

- The key and IV for symmetric-key encryption are derived from a group-shared key $K_g$ and senders public information
- In the one-to-one key exchange phase, after individually computing "Shared Secret" at both sides, there is no key confirmation.
- In the symmetric-key encryption, the sender key ID and recipient key ID are not authenticated.

Some of our attacks are possible with the help of malicious messaging server (E2E adversary). We remark that many messaging application have equipped with an E2E encryption scheme, and the main purpose is to provide a protection against E2E adversary. In addition, we found some attacks that even do not need the help of E2E adversary, which is a severe security flaw.

We have informed our findings to LINE corporation in advance. LINE corporation has confirmed the attacks are valid as long as E2E adversary is involved, while those w/o E2E adversary seem to be thwarted with additional operations not described in the whitepaper, which is hard for us to verify at this point.

### References

1 Tyler Curtiss. Encryption out of LINE Reverse engineering end-to-end encrypted messaging. Ekoparty 2016, 2016.
2 Antonio M. Espinoza, William J. Tolley, Jedidiah R. Crandall, Masashi Crete-Nishihata, and Andrew Hilts. Alice and bob, who the FOCI are they?: Analysis of end-to-end encryption in the LINE messaging application. In USENIX FOCI 17, USENIX Association, 2017.
3 LINE Corporation. LINE Encryption Overview, 2016.

## 3.36 Multiplication Operated Encryption with Trojan Resilience

*Virginie Lallemand (Ruhr-Universität Bochum, DE)*


**Joint work of** Olivier Bronchain, Sebastian Faust, Virginie Lallemand, Gregor Leander, Léo Perrin, François-Xavier Standaert

As most hardware design companies cannot afford having their own foundries, a common strategy consists in outsourcing the production of integrated circuits to external factories. If this solution allows to reduce the production costs, it brings up the problem of trust in the third party. One of the most feared threats in this respect goes under the name of hardware Trojan, defined as a malicious modification of the circuit design. Possible actions of Trojans include moves as devastating as key exfiltration. In this talk, we present a new block cipher construction designed especially to help addressing this problem: our proposal can be implemented using (mostly) untrusted low-cost chips and provides robustness more efficiently than by exploiting secret sharing and multi-party computation on a standard block cipher. Our concrete proposal is called MOE, acronym for "Multiplication Operated Encryption": its round structure only consists in a modular multiplication and a multiplication with a binary matrix. These two operations being linear (with respect to different groups), they allow efficient secret sharing and a reduced hardware cost in comparison to previous solutions. One of our main contribution is the analysis of the cryptographic properties of the modular multiplication, an operation that was used back in the 90s (for the round structure of the ciphers IDEA and MMB for instance) but that to the best of our knowledge was never studied in detail.

## 3.37 Instantiating the Whitened Swap-Or-Not Construction

*Nils Gregor Leander (Ruhr-Universität Bochum, DE)*


**Joint work of** Virginie Lallemand, Gregor Leander, Patrick Neumann, Friedrich Wiemer

We discussed how to instantiate the Whitened Swap-Or-Not Construction by S. Tessaro [1]. We first discussed some inherent limitations and restrictions before showing a first attempt how the framework could be instantiated.

### References

1 Stefano Tessaro. Optimally Secure Block Ciphers from Ideal Primitives. ASIACRYPT (2) 2015: 437–462, Springer, 2015.

### 3.38 Better proofs for rekeying

*Daniel J. Bernstein (University of Illinois – Chicago, US)*


**URL** https://blog.cr.yp.to/20170723-random.html

The current mess of proofs of the cascade (FOCS 1996 Bellare–Canetti–Krawczyk), NMAC (Crypto 1996 Bellare–Canetti–Krawczyk), PRNGs (CCS 2005 Barak–Halevi), and NMAC again (Crypto 2006 Bellare) can be replaced by one simple tight multi-user security proof.

## 4 Panel discussions

### 4.1 Discussion on Mass Surveillance and the Real-World Impact of the Symmetric-Crypto Research Community

*Joan Daemen (Radboud University Nijmegen, NL, and STMicroelectronics – Diegem, BE)*



On Friday morning there was a group discussion open for all seminar participants on a number of questions related to the real-world relevance of the symmetric crypto community and its research. Here a short summary of the outcome of these discussions. We thank Maria Eichlseder for input and taking notes during the discussion. The discussion centered around three themes.

The first theme was education of the general public. All agreed that it is impossible to protect our privacy and security without awareness. This is the case in general and applies specifically to the deployment of cryptography. Whether we, the symmetric cryptographic community, can actually have an impact here, is another thing. A good example is educating the general public about privacy (see mass surveillance, social media, etc.). However, privacy is a very subtle notion and even education on something much simpler as security has failed (see, e.g., how public key cryptography is deployed, or password policies, how people use passwords, etc.). Of course educating developers and policy makers would be easier maybe as they are professionals where a certain level of competence can be expected. Many of us are teaching at universities and there we can make a difference and hope our students will end up in policy-making positions. As a second aspect, the question was raised on what the main messages would be that we want to communicate. Or in other words, is there even a consensus (possible) in the academic community? For example, should companies be allowed to use private data in exchange for services (even after users have agreed to some terms of use)?

The second theme was about the education of protocol designers and programmers. The starting point was that there are many new standards being drafted even now and many repeat the same mistakes over and over again. Often the cryptographic knowledge of people in the standardization committees is very limited. There was discussion where different opinions were expressed and little agreement was reached. What we did agree on is that details of cryptosystems for public use should be made public, and be publicly analyzed. In the past, even public specifications have not always been carefully reviewed. Here the 'provably secure' WPA2, that was recently very badly broken, serves as a good example. As a possible reason for this miserable situation was given that there are 'too many irrelevant standards'. This raised the question: which are the relevant standards that the cryptographic

community should focus on? The point was raised that NIST has usage data that could give us some guidance in this. Another interesting follow-up to this that was raised is that all are encouraged to contribute to updated versions of the Ecrypt CSA document Algorithms, Key Size and Protocol Report.

The third theme was the problem that an activity that is very important for the public and that requires specialized skills and great effort, that of building secure implementations, gives little academic reward. Here it was noted that papers reporting on implementations may be accepted at conferences such as FSE and there are also efforts to create sites with pointers to crypto libraries and tools.

## 4.2 Discussion on Robustness of CAESAR Candidates

*Damian Vizár (EPFL – Lausanne, CH)*



CAESAR (Competition for Authenticated Encryption: Security, Applicability, and Robustness) explicitly names "robustness" as one of the desirable properties that an AE scheme should possess. The call for submissions mentions nonce-misuse resistance, and any candidate may target "any additional security goals and robustness goals that the submitters wish to point out". However, no explicit minimal requirements concerning robustness were *requested* from the CAESAR candidates.

It is not clear whether there is any minimal degree of robustness that any candidate should posses, what kinds of robustness are relevant, and what importance should "robustness" play in the selection of CAESAR finalists. The goal of this discussion is to collect the opinions related to the role of robustness in CAESAR, and to attempt to find a consensus (or a compromise) for the answers to these questions.

### Summary of the Discussion

The initial questions of the discussion were the following:

1. What should be understood under "robustness" in the context of CAESAR?
2. Should there by a degree of "robustness" that is absolutely required from all candidates? (I.e. should there be any hard filtering based on "robustness"?)
3. If "robustness" is required, which particular properties is required, and what degree of resilience is required?

Even though the discussion did not converge to a clear answer to any of the three questions, it did generate a limited number of potential answers to these questions and further useful comments.

**On robustness itself.** As it was pointed out, the term "robustness" is not robust itself. We can mostly agree that informally, robustness means resilience against the improper use of a scheme (or more generally, as Barwell et al. put it "Robustness characterises the ability of a construct to be pushed right to the edge of its intended use case (and possibly beyond)"). Identifying a satisfyingly exact definition in the context of CAESAR seems difficult. These were the comments related to robustness:

- No scheme can be universally robust. There will always be misuse cases that trivially break any scheme (e.g. leaking the secret key in a silly way).

- Currently, robustness is evaluated through formal frameworks (MRAE, OAE, RAE and RUP in CAESAR) which each capture a very precise level of resilience against one or more specific types of misuse.
  - Schemes either claim security in the sense of one of these notions, or they give no guarantees and no information on what happens in case of the related misuse. This could thus be labelled explicit robustness.
- Another possible definition of robustness is that a robust scheme mitigates the damage done by a powerful attack that is outside of its security guarantees. E.g. a nonce based AEAD scheme that only suffers from a non-reusable decryption attack under nonce misuse is more robust (w.r.t. nonce misuse) than one that allows a low-complexity key recovery in the same setting. As this exact level of (in)security is not always advertised by the authors, this could be labelled e.g. implicit robustness.
  - For schemes that make the same claims w.r.t. to explicit robustness, the actual level of (in)security against a strong attack may differ greatly.
- Everyone agrees that side channel resistance is highly desirable. Everyone also agrees that, because side channel protection is platform and implementation-specific, it is best not to include it in this already complicated discussion.
  - It was agreed that the ease of protection against side channel attacks should not be mandatory for all candidates, but should be seen as a strong advantage.
  - Rendering side channel information useless by measures on the protocol level was proposed as a potential research avenue (e.g. using two independent authentication keys to verify firmware updates).

**Required level of robustness.** Especially in this point, no consensus could be reached. However, three general opinions recurred in the discussion:

- **The selection of the finalists should be conservative; the final portfolio should not contain schemes that suffer from devastating attacks, even though these may be outside of their guarantees.** E.g. exclude schemes that allow low-complexity recovery of the secret key or a secret state under nonce misuse. These were the arguments in favour of this opinion:
  - We have to assume that the users of CAESAR recommendations will be inexperienced. They may not understand or may ignore the usage conditions of the finalists. "The good engineers will not need the portfolio."
  - There are bound to be cases of misuse, and we should try to mitigate the damage, at least for those kinds of misuse that we understand.
  - There are bound to be cases of (nonce) misuse, in which the devastating attacks may undermine the credibility of the symmetric cryptography. This will be the opposite effect of what CAESAR aims for.
  - The current pool of candidates contain schemes that do not suffer from devastating attacks, why not take those?
- **There should be no default level of robustness required from the candidates. We should not eliminate candidates based on a default robustness criterion.** These were the arguments in favour of this opinion:
  - It is enough that the finalists come with simple labels that clearly state what must and what must not be done to preserve security. It is the responsibility of the users to follow the (simple) instructions.
  - There are simple ways of making sure that the relevant misuse never occurs (e.g. device-specific prefixes in nonces).

  - This was not demanded at the beginning, or during the competition. We should not introduce improvements over AES-GCM's robustness now.
  - We already have the use cases to take care of this. In particular, this is not the primary concern in the "high-performance applications" use case.
  - We cannot thwart every kind of misuse (e.g. using key as a random IV), thus we should not make particular forms of robustness compulsory.
- **Something in between.** There were two major proposals of the in-between kind:
  - **No default robustness requirement. Take into account the cryptanalysis, consider each case individually. Use the cryptanalysis to break ties.** The idea of using the results on exact (in)security of CAESAR candidates for breaking ties between similar schemes in the final decision process seemed to be generally well accepted.
  - **No default robustness requirement. When issuing final portfolio, give 2 kinds of labels to all finalists: (1) "regular schemes" and (2) "experts-only schemes" (or "brittle schemes").** The regular schemes would be those with no devastating low complexity nonce reuse attacks or nonce respecting birthday attacks. These would be recommended for a common user. The expert-only schemes would get a warning of dire consequences in case of misuse and their brittleness.

**The most desirable forms of robustness.** This point was not addressed in much detail, as the discussion focused mostly on the issue of having or not having default robustness requirement. However, most of the examples, counter examples and comments worked with nonce reuse.

## Participants


- Frederik Armknecht
Universität Mannheim, DE
- Tomer Ashur
KU Leuven, BE
- Christof Beierle
Ruhr-Universität Bochum, DE
- Daniel J. Bernstein
University of Illinois – Chicago, US
- Eli Biham
Technion – Haifa, IL
- Alex Biryukov
University of Luxembourg, LU
- Anne Canteaut
INRIA – Paris, FR
- Joan Daemen
Radboud University Nijmegen, NL, and STMicroelectronics – Diegem, BE
- Itai Dinur
Ben Gurion University – Beer Sheva, IL
- Christoph Dobraunig
TU Graz, AT
- Orr Dunkelman
University of Haifa, IL
- Maria Eichlseder
TU Graz, AT
- Henri Gilbert
ANSSI – Paris, FR
- Tetsu Iwata
Nagoya University, JP
- Jérémy Jean
ANSSI – Paris, FR
- Dmitry Khovratovich
University of Luxembourg, LU
- Stefan Kölbl
Technical University of Denmark – Lyngby, DK
- Virginie Lallemand
Ruhr-Universität Bochum, DE
- Tanja Lange
TU Eindhoven, NL
- Nils Gregor Leander
Ruhr-Universität Bochum, DE
- Gaëtan Leurent
INRIA – Paris, FR
- Stefan Lucks
Bauhaus-Universität Weimar, DE
- Willi Meier
FH Nordwestschweiz – Windisch, CH
- Bart Mennink
Radboud University Nijmegen, NL
- Vasily Mikhalev
Universität Mannheim, DE
- Kazuhiko Minematsu
NEC – Kawasaki, JP
- Nicky Mouha
NIST – Gaithersburg, US
- Mridul Nandi
Indian Statistical Institute – Kolkata, IN
- Maria Naya-Plasencia
INRIA – Paris, FR
- Kaisa Nyberg
Aalto University, FI
- Stav Perle
Technion – Haifa, IL
- Léo Paul Perrin
INRIA – Paris, FR
- Thomas Peyrin
Nanyang TU – Singapore, SG
- Christian Rechberger
TU Graz, AT
- Arnab Roy
University of Bristol, GB
- Yu Sasaki
NTT – Tokyo, JP
- Yannick Seurin
ANSSI – Paris, FR
- Adi Shamir
Weizmann Institute – Rehovot, IL
- Marc Stevens
CWI – Amsterdam, NL
- Stefano Tessaro
University of California – Santa Barbara, US
- Yosuke Todo
NTT – Tokyo, JP
- Gilles Van Assche
STMicroelectronics – Diegem, BE
- Damian Vizár
EPFL – Lausanne, CH
- Meiqin Wang
Shandong University – Jinan, CN
- Kan Yasuda
NTT – Tokyo, JP


Report from Dagstuhl Seminar 18031

# Personalized Multiobjective Optimization: An Analytics Perspective

**Edited by**

**Kathrin Klamroth[1], Joshua D. Knowles[2], Günter Rudolph[3], and Margaret M. Wiecek[4]**

1 **Universität Wuppertal, DE,** klamroth@math.uni-wuppertal.de
2 **University of Birmingham, GB,** j.knowles@cs.bham.ac.uk
3 **Technische Universität Dortmund, DE,** guenter.rudolph@tu-dortmund.de
4 **Clemson University, US,** wmalgor@clemson.edu

## Abstract

The Dagstuhl Seminar 18031 Personalization in Multiobjective Optimization: An Analytics Perspective carried on a series of five previous Dagstuhl Seminars (04461, 06501, 09041, 12041 and 15031) that were focused on Multiobjective Optimization. The continuing goal of this series is to strengthen the links between the Evolutionary Multiobjective Optimization (EMO) and the Multiple Criteria Decision Making (MCDM) communities, two of the largest communities concerned with multiobjective optimization today. Personalization in Multiobjective Optimization, the topic of this seminar, was motivated by the scientific challenges generated by personalization, mass customization, and mass data, and thus crosslinks application challenges with research domains integrating all aspects of EMO and MCDM. The outcome of the seminar was a new perspective on the opportunities as well as the research requirements for multiobjective optimization in the thriving fields of data analytics and personalization. Several multi-disciplinary research projects and new collaborations were initiated during the seminar, further interlacing the two communities of EMO and MCDM.

**Seminar** January 14–19, 2018 – http://www.dagstuhl.de/18031
**2012 ACM Subject Classification** Mathematics of computing → Mathematical optimization, Applied computing → Operations research, Computing methodologies → Artificial intelligence
**Keywords and phrases** multiple criteria decision making, evolutionary multiobjective optimization
**Digital Object Identifier** 10.4230/DagRep.8.1.33
**Edited in cooperation with** Richard Allmendinger

## 1 Executive Summary

*Kathrin Klamroth*
*Joshua D. Knowles*
*Günter Rudolph*
*Margaret M. Wiecek*



The topic of the seminar, Personalization in Multiobjective Optimization, was motivated by ongoing changes in many areas of human activity. In particular, personalization, mass


Personalized Multiobjective Optimization: An Analytics Perspective, *Dagstuhl Reports*, Vol. 8, Issue 01, pp. 33–99
Editors: Kathrin Klamroth, Joshua D. Knowles, Günter Rudolph, and Margaret M. Wiecek
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

**Table 1** Working groups (WGs) crosslinking application challenges (rows) with research domains (columns). WG 1: Preference uncertainty quantification; WG 2: Personalization and customization of decision support; WG 3: Invariant rule extraction; WG 4: Complex networks and MCDA; WG 5: Metamodelling for interactive optimization.

| | Modelling | Preferences | Algorithms |
|---|---|---|---|
| Platform design and product lines | WG3, WG5 | WG1, WG3 | WG3 |
| Responsive and online personalization | WG2, WG5 | WG1, WG2 | WG2 |
| Complex networks of decision makers | WG4, WG5 | WG1, WG4 | WG4 |

customization, and mass data have become essential in current business and engineering operations creating new challenges for academic and research communities. In the seminar, the EMO and MCDM communities, including junior and senior academic researchers as well as industry representatives, took an effort to jointly address the ongoing changes in the real-world with multiobjective optimization.

The purpose of multiobjective optimization is to develop methods that can solve problems having a number of (conflicting) optimization criteria and constraints, providing a multitude of solution alternatives, rather than pursuing only one "optimal" solution. In this aim the field has been highly successful: its methods have a track record of improving decision making across a broad swath of applications, indeed wherever there are conflicting goals or objectives. Yet, multiobjective optimization has so far focused almost exclusively on serving a single "decision maker", providing solutions merely as potential (not actual) alternatives. In order to fulfill the demanding aims of mass-customization, product/service variation and personalization we see today in areas such as engineering, planning, operations, investment, media and Web services, and healthcare, new and innovative approaches are needed. This seminar took the first steps towards this goal by bringing together leading specialists in EMO and MCDM.

Personalization in multiobjective optimization as the main theme of the seminar has focused around three **application challenges** which are highly characteristic for real-world decision making and represent different ways that personalization is needed or delivered in an optimization setting. These were (i) Platform design and product lines, (ii) Responsive and online personalization, and (iii) Complex networks of decision makers. These three application challenges were crosslinked with three **research domains** that constitute the methodological core of multiobjective optimization and have been the foundation for the discussions at the previous Dagstuhl seminars. These were (1) Model building, (2) Preference modelling, and (3) Algorithm design and efficiency.

During the seminar, we formed five multi-disciplinary working groups (WGs) to implement the crosslinking between these application challenges and research domains, see Table 1. Each working group was focused on an application challenge (a row in Table 1; WGs 2, 3 and 4) or a research domain (a column in Table 1; WGs 1 and 5), all taking specific perspectives on the respective topics.

The program was updated on a daily basis to maintain flexibility in balancing time slots for talks, discussions, and working groups. The working groups were established on the first day in an open and highly interactive discussion. The program included several opportunities to report back from the working groups in order to establish further links and allow for adaptations and feedback. Some of the working groups split into subgroups and rejoined later in order to focus more strongly on different aspects of the topics considered. Abstracts of the talks and extended abstracts of the working groups can be found in subsequent chapters

of this report. Further notable events during the week included: (i) a hike on Wednesday afternoon with some sunshine (despite the quite terrible weather during the rest of the week), (ii) an announcements session allowing us to share details of upcoming events in our research community, and (iii) a wine and cheese party made possible by the support of the ITWM Kaiserslautern, represented by Karl-Heinz Küfer.

## Outcomes

Fourteen topical presentations were complemented by discussions in five working groups, covering the main themes of the seminar. The outcomes of each of the working groups can be seen in the sequel. Extended versions of their findings will be submitted to a Special Issue on "Personalization in Multiobjective Optimization: An Analytics Perspective" of the *Journal of Multicriteria Decision Analysis*, edited by Theo Stewart, that is guest edited by the organizers of this seminar. The submission deadline is July 31, 2018, and several working groups plan to submit extended versions of their reports to this special issue.

The seminar was highly productive, very lively and full of discussions, and has thus further strengthened the interaction between the EMO and MCDM communities. We expect that the seminar will initiate a new research domain interrelating multiobjective optimization and personalization, as it similarly has happened after the previous seminars in this series.

## Acknowledgments

A huge thank you to the Dagstuhl office and its very helpful and patient staff; many thanks to the organizers of the previous seminars in the series for the initiative and continuing advice; and many thanks to all the participants, who contributed in so many different ways to make this week a success. In the appendix, we also give special thanks to Joshua Knowles as he steps down from the organizer role.

## 2 Table of Contents

# 3 Overview of Talks

## 3.1 Industrial applications of multicriteria decision support systems

*Karl Heinz Küfer (Fraunhofer ITWM – Kaiserslautern, DE)*



Most decisions in life are compromises: several objectives, most often arising from the four families cost, quality, time or environmental impact, have to be balanced. Decision making is rarely straight-forward because one cannot have best possible values for all of these goals simultaneously as they are at least partially in conflict. Many decision makers are reluctant to introduce decision support tools that directly show what the possible freedom of choice or inherent restrictions of the problems are. They often do not want to defend personal preferences or biases in decision rounds, which would become obvious by showing options and limitations in a transparent way. Others are in sorrows concerning the profile of or even their jobs. The talk will demonstrate and discuss examples of decision support tools in medical therapy planning, chemical process engineering and in the layout of renewable energy facilities, all of them in industrial practice for five or more years. Special attention is paid to the reception of such concepts in the companies and their impact if successfully implemented.

## 3.2 Culturally tailored multicriteria product design using crowdsourcing

*Georges Fadel (Clemson University – Clemson, US)*


**Joint work of** Georges Fadel, Ivan Mata, Mo Chen, Paolo Guarneri, Manh Tien Nguyen
**Main reference** Ivan Mata, Georges Fadel, Anthony Garland, Winfried Zanker: "Affordance based interactive genetic algorithm (ABIGA)", Design Science, Vol. 4, E5, 2018.
**URL** https://doi.org/10.1017/dsj.2017.30

The presentation describes an approach to involve crowds of users in the evolution of the design of a product by having them provide feedback to a tailored interactive multi-objective archive based micro- genetic algorithm. Affordances are defined as perceived opportunities for action, for instance, a ladder affords elevating the user and a glass affords containing a liquid. The users grade perceived affordances of a product and these are the criteria that the GA uses to evolve the shape of a product. The algorithm has multiple archives that store culturally biased solutions and use them in the evolution of solutions. After a number of generations, the designer can extract from the stored data which physical parameters affect specific affordances in the view of the users. The users will eventually be able to suggest additional affordances, and the designer would have to accept or not to add such a criterion to the system, and have possibly the designs evolve differently. A set of non-dominated solutions is then available to the designer to choose from. The system can be used by an individual to personalize a solution, or by a crowd to evolve the solution towards a more satisfycing solution to the group.

## 3.3 Metamodeling approaches for multiobjective optimization

*Kalyanmoy Deb (Michigan State University – East Lansing, US)*



In multiobjective optimization, every objective function must be approximated with a suitable metamodel, particularly when a solution evaluation is computationally expensive. One straightforward approach is to model every objective function separately, but a number of other approaches are possible and may be more effective. In this talk, we proposed a taxonomy of different metamodeling frameworks and presented our recent results of each framework on multiple test problems. This research is motivated by practice and opens up a number of avenues for new research and application. Some of the methods highlighted are: (i) Specific metamodeling approaches (Kriging, RBF, or others) and their choice for every objective and constraint function, (ii) possible switching methods from one framework to another with iterations, (iii) possible other selection methods for metamodeling based on EMO methodologies, and (iv) possible use of trust region methods along with metamodeling approaches. Results on an industrial design problem was presented.

## 3.4 Representations: Do they have potential for customer choice?

*Serpil Sayın (Koc University – Istanbul, TR)*


**Joint work of** Serpil Sayın, Gokhan Kirlik
**Main reference** Gokhan Kirlik, Serpil Sayın: "Bilevel programming for generating discrete representations in multiobjective optimization", Math. Program., Vol. 169(2), pp. 585–604, 2018.
**URL** http://dx.doi.org/10.1007/s10107-017-1149-0

Representations are subsets of nondominated sets that are expected to serve in the capacity of the original set. Finding representations makes most sense when the latter set is computationally difficult to obtain or practically difficult to explore. In recent years, there have been a number of studies that focused on delivering representations for multiobjective optimization problems. Some of these studies propose measures of quality to assess how well a representation or an approximation mimics the original set. These studies are mostly set in environments where finding the entire nondominated set is computationally challenging. Therefore they have not been discussed from the perspective of representing sets when all alternatives are explicitly available.

One problem in online retailing is presenting the items in a category to a potential customer. In most cases, the category contains a large selection of items. The user usually has a number of ways to customize the way she explores the category. For instance, filters may help limit values of interest for some relevant criteria. There may be choices offered to sort the items with respect to price, popularity, etc. I would like to ask the question if it is possible to design a new way of presenting a category to a customer based on what we know about representing nondominated sets. This would call for casting a customer's product choice problem as a multiple criteria one and delivering alternative mechanisms of navigating the category.

This discussion relates to the application challenge responsive and online personalization as well as representations in research domain.

## 3.5 Modelling complex networks of decision makers: An analytical sociology perspective

*Robin Purshouse (University of Sheffield – Sheffield, GB)*


**Joint work of** Robin C. Purshouse, Shaul Salomon, Gideon Avigad, Peter J. Fleming
**Main reference** Shaul Salomon, Robin C. Purshouse, Gideon Avigad, Peter J. Fleming: "An Evolutionary Approach to Active Robust Multiobjective Optimisation", in Proc. of the Evolutionary Multi-Criterion Optimization - 8th International Conference, EMO 2015, Guimarães, Portugal, March 29 -April 1, 2015. Proceedings, Part II, Lecture Notes in Computer Science, Vol. 9019, pp. 141–155, Springer, 2015.
**URL** https://doi.org/10.1007/978-3-319-15892-1_10

Designers and planners who provide solutions for mass-markets and communities wish to understand how individuals in those markets and communities make choices about how they use, customise or reject those solutions. For example, a powertrain designer with fleet-level emissions and durability objectives wants to understand the different ways in which owners might operate a plug-in hybrid vehicle; a government planner with community-level health and revenue objectives wants to understand how citizens might choose to exploit a subsidised recreational facility. Whilst formulation of the higher-level multi-objective decision problem facing designers and planners has been addressed many times by researchers, far less attention has been paid to the, typically repeated, lower-level multi-objective decision problem faced by users, or to the interaction between these levels. Decisions at the lower-level are embedded within a complex socio-technical context, in which interactions between individuals can play a key role in how decisions are made and changed over time. This talk will introduce the framework of analytical sociology, pioneered by the Swedish sociologist Peter Hedström, as a means of modelling mass-customisation decision problems. Analytical sociology is a theory-based approach in which individual behaviours are driven by specified causal mechanisms. The talk will describe the three types of mechanism captured by the framework – situational, individual action, and transformational – and highlight the potential role of the designer and planner in shaping the decisions of heterogeneous individuals in mass-markets and communities.

## 3.6 Data-driven automatic design of multi-objective optimizers

*Manuel López-Ibáñez (University of Manchester – Manchester, GB)*



Recent work is increasingly showing that, given a library of good algorithmic components, automatically designed algorithms consistently outperform human-designed ones, even for thoroughly researched benchmark problems [1, 2, 3, 4]. The benefits of automated algorithm design rapidly increase for more complex and less studied problems, where the intuitions of human experts often fail. The transition from an expert-driven human-intensive design methodology to a data-driven CPU-intensive one also leads to the production of large amounts of data about the performance of algorithmic components. Despite some initial work in single-objective optimization and machine learning [5], it is still an open question how to use and analyze this data to gain insights about algorithmic components applied to multi-objective problems. Moreover, the transition to an automated design methodology

raises questions about performance metrics, the identification of equivalent and alternative algorithmic components, and the role of the decision-maker; questions that are particularly relevant in a multi-objective context.

### References


1 H. H. Hoos. Programming by optimization. *Communications of the ACM*, 55(2):70–80, 2012.

2 A. R. KhudaBukhsh, L. Xu, H. H. Hoos, and K. Leyton-Brown. SATenstein: Automatically Building Local Search SAT Solvers from Components. *Artificial Intelligence*, 232:20–42, 2016.

3 M. López-Ibáñez and T. Stützle. The Automatic Design of Multi-Objective Ant Colony Optimization Algorithms. *IEEE Transactions on Evolutionary Computation*, 16(6):861–875, 2012.

4 L. C. T. Bezerra, M. López-Ibáñez , and T. Stützle. Automatic Component-Wise Design of Multi-Objective Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation*, 20(3):403–417, 2016.

5 F. Hutter, H. H. Hoos, and K. Leyton-Brown. An Efficient Approach for Assessing Hyperparameter Importance. In *Proceedings of the 31th International Conference on Machine Learning*, volume 32, pages 754–762, 2014.


## 3.7 Maximizing the probability of consensus in group decision making

*Michael Emmerich (Leiden University – Leiden, NL)*


**Joint work of** Michael Emmerich, Andre Deutz, Iryna Yevseyeva

Consider the scenario of selecting a portfolio of $k$ alternative solutions from a set of $n >> k$ solutions. A moderator presents $k$ solutions to a board of decision makers. The goal is to maximize the probability that the decision makers achieve consensus about at least one solution in the portfolio. In advance, decision makers formulated desirability functions for the objectives of concern – ranging from 0 (not acceptable) to 1 (fully satisfactory). Moreover, correlations between objectives may be formulated using a dependence graph. The analysis shows that the computation of the probability of consensus is related to specific integrals over the dominated space, which reduces to the hypervolume indicator after coordinate transformation in case of independent objectives. The problem of veto by overdemanding decision makers is discussed we propose a possible remedy by replacing the probability by higher momenta of the joint acceptance probability distribution.

## 3.8 Decision analytics with multiobjective optimization and a case in inventory management

*Kaisa Miettinen (University of Jyväskylä – Jyväskylä, FI)*


**Joint work of** Kaisa Miettinen, Juha Sipilä, Risto Heikkinen, Vesa Ojalehto
**URL** http://www.jyu.fi/demo

Thanks to digitalization, we have access to various types of data and must decide how to make the most of the data. We can use descriptive or predictive analytics but to make recommendations and informed decisions based on the data, we need prescriptive or decision analytics. If the problems contain multiple conflicting objectives, multiobjective optimization are to be applied.

We introduce the new thematic research area at the University of Jyväskylä called Decision Analytics utilizing Causal Models and Multiobjective Optimization (DEMO). The objective of DEMO is to develop elements of a seamless chain from data to decision support.

Lot sizing is an example of a data-driven optimization problem. It is important in production planning and inventory management, where a decision maker needs support, in particular, when the demand is stochastic. We consider the lot sizing problem of a Finnish production company and formulate four conflicting objectives. We solve it with two interactive multiobjective optimization methods. In interactive methods, a decision maker directs the search for the best balance between the conflicting objectives by providing preference information. In this way, (s)he can learn about what kind of solutions are available for the problem and also learn about the feasibility of one's preferences.

In the case considered, the decision maker found it useful to switch the method during the solution process. The results of this data-driven interactive multiobjective optimization approach are encouraging and demonstrate the practical value of decision analytics.

## 3.9 Actively learning a mapping for personalisation

*Jürgen Branke (University of Warwick – Warwick, GB)*



This talk tackles the problem of efficiently collecting data to learn a classifier, or mapping, from each user to the best personalisation, where users are described by continuous features and there is a finite set of personalisation options to choose from. An example would be online advertisements, where we want to learn the best possible advertisement and advertisement format for each user. We propose a fully sequential information collection policy based on Bayesian statistics and Gaussian Process models. In each step, they myopically allocate to the user the advertisement that promises the highest value of information collected.

## 3.10 The NEMO framework for EMO: Learning value functions from pairwise comparisons

*Roman Słowiński (Poznan University of Technology – Poznan, PL)*


**Joint work of** Roman Słowiński, Jürgen Branke, Salvatore Corrente, Salvatore Greco

Some years ago, we have proposed the NEMO framework to enhance multi-objective evolutionary algorithms by pairwise preference elicitation during the optimisation, allowing the algorithm to converge more quickly to the most relevant region of the Pareto front. The framework is based on Robust Ordinal Regression. Over the years, several variations have been developed, with different user preference models (linear, additive, Choquet-integral value functions) and different ways of integrating this information into evolutionary algorithms (as a surrogate fitness function, or by enriching the dominance relation). This presentation will provide an overview of the developments in this area.

## 3.11 Uncertainty quantification on Pareto fronts

*Mickaël Binois (University of Chicago – Chicago, US)*



In this short presentation, we review methods to approximate Pareto fronts in the case of expensive, possibly noisy, blackbox objective functions. We concentrate on methods involving Gaussian processes, which provide uncertainty quantification on the estimated Pareto front. Variations on the modeling include re-interpolation and nugget estimation, while the uncertainty is estimated from sampling, random closed sets or bootstrap.

## 3.12 Innovization: Unveiling invariant rules from non-dominated solutions for knowledge discovery and faster convergence

*Abhinav Gaur (Michigan State University – East Lansing, US)*



Multi-objective optimization (MOO) problems lend themselves to not one but a set of optimal solutions also called Pareto-optimal (PO) solutions. Such PO solutions carry information on patterns that make these solutions concurrently optimal for multiple objectives/Customer preferences. Discovering such patterns from the PO solutions is called 'Innovization' or innovation through optimization. Some of the uses of carrying out an Innovization exercize are discovering principles that makes certain solutions PO for a MOO problem, automatically discovering optimization heuristics for a problem and, expediting black box MOO algorithms. In the context of "Personalized MOO", the concepts in Innovization, Higher Level Innovization, Lower Level Innovization, Temporal Innovization have direct applications. For example, temporal Innovization can help us discover principles that govern how preferences of a class of customer have changed over time. Lower level innovization can help us discover preferences

of customers whose preferences lie at part of the PO front. Higher level Innovization can help us discover principles that govern the customer preferences as certain problem parameters are changed, and so on. Hence, the Innovization idea seems to be very relevant to the problem of studying "Personalized Multi Objective Optimization".

## 3.13 Compressed data structures for bi-objective 0,1-knapsack problems

*José Rui Figueira (IST – Lisbon, PT)*



Solving multi-objective combinatorial optimization problems to optimality is a computationally expensive task. The development of implicit enumeration approaches that efficiently explore certain properties of these problems has been the main focus of recent research. This article proposes algorithmic techniques that extend and empirically improve the memory usage of a dynamic programming algorithm for computing the set of efficient solutions both in the objective space and in the decision space for the bi-objective knapsack problem. An in-depth experimental analysis provides further information about the performance of these techniques with respect to the tradeoff between CPU time and memory usage.

## 3.14 Recent algorithmic progress in multiobjective (combinatorial) optimization

*Andrzej Jaszkiewicz (Poznan University of Technology – Poznan, PL)*



Despite of many years of research in the area of multiobjective evolutionary algorithms and more generally multiobjective metaheuristics many real-life multiobjective problems, in particular combinatorial problems, constitute a serious challenge for existing methods. Recently an important progress has been made in the algorithmic toolbox of multioobjective optimization. Some of the new algorithms are focused on the combinatorial optimization, but many are more generally applicable. Some of the recently proposed or improved algorithms are:

- ND-Tree data structure and algorithm for the dynamic non-dominance problem [1]. ND-Tree allows for very efficient update of even large Pareto archives. It allows multiobjective evolutionary algorithms and other metaheuristics to store large sets of potentially Pareto-optimal solutions without loss of efficiency. ND-Tree can also be applied to efficiently solve the non-dominated sorting problem often used in evolutionary algorithms.
- Many-objective Pareto Local Search (MPLS) [2]. Pareto Local Search proved to be a very effective tool in the case of the bi-objective combinatorial optimization and it was used in a number of the state-of-the-art algorithms for problems of this kind. On the other hand, the standard Pareto Local Search algorithm becomes very inefficient for problems with more than two objectives. Many-Objective Pareto Local Search algorithm uses three new mechanisms to preserve the effectiveness of PLS in many-objective case.

The new mechanisms are: the efficient update of large Pareto archives with ND-Tree data structure, a new mechanism for the selection of the promising solutions for the neighborhood exploration, and a partial exploration of the neighborhoods.

- New efficient algorithms for calculating the exact hypervolume of the space dominated by a set of d-dimensional points. This value is often used as the quality indicator in the multiobjective evolutionary algorithms and other metaheuristics and the efficiency of calculating this indicator is of crucial importance especially in the case of large sets or many dimensional objective spaces. Recently significant improvements have been obtained in algorithms for calculating this indicator [3, 4, 5, 6, 7, 8]. They allow not only to speed-up computational experiments but also to use hypervolume within multiobjective algorithms, e.g. to guide the search or to define stopping conditions.

### References


1 A. Jaszkiewicz and T. Lust. ND-Tree-based update: a Fast Algorithm for the Dynamic Non-Dominance Problem. *ArXiv e-prints*, arXiv:1603.04798, 2016.

2 A. Jaszkiewicz. Many-Objective Pareto Local Search. *ArXiv e-prints*, arXiv: 1707.07899, 2017.

3 A. Jaszkiewicz. Improved quick hypervolume algorithm. *Computers & Operations Research*, 90:72–83, 2018.

4 L. M. S. Russo and A. P. Francisco. Quick hypervolume. *IEEE Transactions on Evolutionary Computation*, 18(4):481–502, 2014.

5 L. M .S. Russo and A. P. Francisco. Extending quick hypervolume. *Journal of Heuristics*, 22(3):245–271, 2016.

6 R. Lacour, K. Klamroth, and C. M. Fonseca. A box decomposition algorithm to compute the hypervolume indicator. *Computers & Operations Research*, 79:347–360, 2017.

7 L. While, L. Bradstreet, and L. Barone. A fast way of calculating exact hypervolumes. *IEEE Transactions on Evolutionary Computation*, 16(1):86–95, 2012

8 W. Cox and L. While. Improving the IWFG Algorithm For Calculating Incremental Hypervolume. *IEEE Congress on Evolutionary Computation*, pages 39690–3976, 2016.


# 4 Working Groups (WGs)

## 4.1 Multi-criteria decision making under performance and preference uncertainty (WG1)

*Mickaël Binois, Jürgen Branke, Alexander Engau, Carlos M. Fonseca, Salvatore Greco, Miłosz Kadziński, Kathrin Klamroth, Sanaz Mostaghim, Patrick Reed, and Roman Słowiński*



**Abstract.** We propose a novel methodology for interactive multi-objective optimization taking into account imprecision, ill-determination and uncertainty referring to both, the technical aspects determining evaluations of solutions by objective functions and the subjective aspects related to the preferences of the decision maker. With this aim, we consider a probability distribution on the space of the objective functions and a probability distribution on the space of the utility functions representing preferences of the decision maker. On the basis of these two probability distributions, without loss of generality supposed to be independent, one can compute a multi-criteria expected utility with a corresponding standard

deviation, that permit to assess a quality of each proposed solution. One can also compute an average multi-criteria expected utility and a related standard deviation for a set of solutions, which permit to assess a quality of a population of solutions. This feature can be useful in evolutionary multi-objective optimization algorithms to compare populations of solutions in successive iterations.

### 4.1.1 Introduction

This paper summarizes the work of the *Preference Uncertainty Quantification* working group at the Dagstuhl seminar 18031 "Personalized Multi-objective Programming: An Analytics Perspective" that took place in Schloss Dagstuhl – Leibniz Center for Informatics - on January 14–19, 2018.

### 4.1.2 Uncertainties

When dealing with multi-objective optimization problems, the decision makers (DMs), and the analysts helping them to solve these problems, are confronted in their reasoning with some uncertainties that are inherent to two kinds of "imperfect" information (see [2] and [3]):

1. Information about the preferences of DMs is always partial and ill-defined. Even more, complete preferences do not exist a priori in DMs' mind, because they evolve in the decision aiding process in interaction with an analyst. The preferences are formed in a constructive learning process in which DMs get a conviction that the most preferred solution has been reached for a given problem statement.
2. Information about consequences of considered solutions usually depend on hardly measurable or random variables. This makes that, in general, the evaluation of solutions with respect to different criteria is imprecise or uncertain.

Therefore, there is a need to take into account these two sources of uncertainty in an interactive multi-objective optimization process. A first consideration of this problem, but taking into account only uncertainty related to utility functions, has been proposed in [4].

### 4.1.3 Problem formulation and basic notation

The multi-objective optimization process presented in this paper is formally represented as a multi-objective programming problem under performance and preference uncertainty as follows. Let $X \subset R^n$ be an $n$-dimensional set of feasible decisions (or solutions, designs, alternatives, etc.) Let $f\colon R^n \to R^m$ be an $m$-dimensional vector, called objective function, that maps each decision $x \in X$ to a corresponding consequence or performance vector $y = f(x)$. To model performance uncertainty, we assume that each objective function $f = (f_1, f_2, \ldots, f_m)$ is a random element of some (for now: a priori) given set $\mathcal{F}$ of cardinality $k$, i.e., $\mathcal{F} = \{f^1, f^2, \ldots, f^k\}$ with random outputs $y^i = (y^i_1, y^i_2, \ldots, y^i_m)$ for each $i \in \{1, 2, \ldots, k\}$. In other words, for each $i \in \{1, 2, \ldots, k\}$, the vector function $f^i = (f^i_1, f^i_2, \ldots, f^i_m)$ is one realization of the random objective function $f$.

Moreover, under the additional assumption that this uncertainty is stochastic in nature, we can assign or estimate a stochastic probability vector $p = (p_1, p_2, \ldots, p_k)$ with $\sum_{i=1}^{k} p_i = 1$ and with the interpretation that $\Pr[f = f^i] = \Pr[y = y^i] = p_i$ for each $i \in \{1, 2, \ldots, k\}$. In this way, we have defined a discrete probability distribution on the space of values taken by the objective function. Obviously, one can consider a generic probability distribution, not necessarily a discrete one. For a scheme of this setting, see the conceptual relationship between *technical information about the performance* and *conjoint probability distribution on values of the objective function* in Figure 1 on the top.

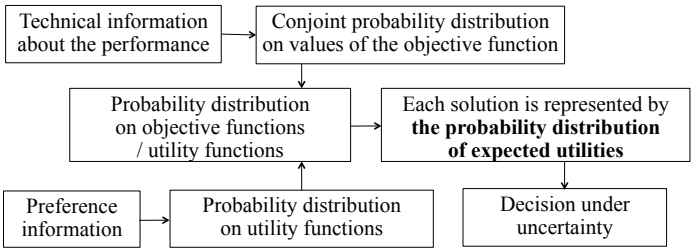



**Figure 1** Main idea underlying the proposed methodology.

Similarly, we can describe the uncertainty about preferences of the DM, considering a utility function $u : R^m \to R$, such that $y \mapsto z = u(y)$. Again, $u$ is considered to be an element of a set $\mathcal{U} = \{u^1, \ldots, u^\ell\}$, interpreted as a set of possible realizations of an uncertain utility function. Each utility function $u^j \in \mathcal{U}$ has a probability $\Pr[u = u^j] = q_j$, $j = 1, \ldots, \ell$. This is marked in Figure 1 as *preference information* and *probability distribution of utility function*.

### A simple example

Consider a simple example, with $n = 2$ and $X = [0, 1]^2$, so that the decision input to the objective functions is a vector $x = (x_1, x_2)$ composed of two decision variables.

**Performance uncertainty.** Let us measure the performance of $x$ in two dimensions, i.e., $m = 2$, so that $f\colon R^2 \to R^2$ with $f = (f_1, f_2)$ for each objective realization. Moreover, consider $k = 3$ uncertain realizations of the objective function, denoted by $\mathcal{F} = \{f^1, f^2, f^3\}$, with probabilities $p = (p_1, p_2, p_3) = (0.5, 0.2, 0.3)$, and taking the following form:

$$f^1(x) := (f_1^1(x_1, x_2), f_2^1(x_1, x_2)) = (x_1, x_2)$$
$$f^2(x) := (f_1^2(x_1, x_2), f_2^2(x_1, x_2)) = (\sqrt{x_1}, \sqrt[3]{x_2})$$
$$f^3(x) := (f_1^3(x_1, x_2), f_2^3(x_1, x_2)) = (x_1^2, x_2^3).$$

**Note:** Alternatively, supposing that the values taken by the objective function in each realization depend on the value taken on a basic reference realization (for example the mean value in case of an estimation through a Bayesian process) one can define the performance set $Y := \{f(x)\colon x \in X\} \subset R^m$ and then use a transformation $\phi^h\colon R^m \to R^m$ for each possible realization $h = 1, \ldots, k$, so that for each $y = f(x) \in Y$ we can also write $\phi^h(y) = \phi^h(y_1, y_2)$ or $\phi^h(f_1(x), f_2(x)) = (f_1^h(x), f_2^h(x))$. For instance, in the considered example, we can take as a basic reference realization $f^1(x) = f^1(x_1, x_2) = (f_1(x_1, x_2), f_2(x_1, x_2)) = (y_1, y_2) = (x_1, x_2)$, and for each realization $h = 1, 2, 3$, suppose:

- $\phi^1(y) = \phi^1(y_1, y_2) = (y_1, y_2)$
  or $\phi^1(f(x)) = \phi^1(f_1(x), f_2(x)) = (f_1^1(x), f_2^1(x)) = (f_1(x), f_2(x))$,
- $\phi^2(y) = \phi^2(y_1, y_2) = (\sqrt{y_1}, \sqrt[3]{y_2})$
  or $\phi^2(f(x)) = \phi^2(f_1(x), f_2(x)) = (f_1^2(x), f_2^2(x)) = (\sqrt{f_1(x)}, \sqrt[3]{f_2(x)})$,
- $\phi^3(y) = \phi^3(y_1, y_2) = ((y_1)^2, (y_2)^3)$
  or $\phi^3(f(x)) = \phi^2(f_1(x), f_2(x)) = (f_1^3(x), f_2^3(x)) = ((f_1(x))^2, (f_2(x))^3)$.

**Preference uncertainty.** Suppose that we have a probability distribution on a set of $\ell = 4$ utility functions describing the preference information as follows:

$$\begin{aligned} q_1 = 0.4 &: u^1(y) = 0.3y_1 + 0.7y_2, \\ q_2 = 0.3 &: u^2(y) = 0.5y_1 + 0.5y_2, \\ q_3 = 0.2 &: u^3(y) = 0.8y_1 + 0.2y_2, \\ q_4 = 0.1 &: u^4(y) = 0.9y_1 + 0.1y_2, \end{aligned}$$

where $q_1, q_2, q_3, q_4$ are probabilities of realization of these utility functions.

**Expected utility and variance of a single solution.** In the following, we assume that the probability distributions of performance information and utility functions are independent from each other. Therefore, the joint probability distribution on the product space $\mathcal{F} \times \mathcal{U}$ assigns to each pair $(f^i, u^j)$ the probability $\pi_{ij} = p_i \cdot q_j$ shown in the following matrix:

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{21} & \pi_{31} \\ \pi_{12} & \pi_{22} & \pi_{32} \\ \pi_{13} & \pi_{23} & \pi_{33} \\ \pi_{14} & \pi_{24} & \pi_{34} \end{bmatrix}^T = \begin{bmatrix} 0.20 & 0.08 & 0.12 \\ 0.15 & 0.06 & 0.09 \\ 0.10 & 0.04 & 0.06 \\ 0.05 & 0.02 & 0.03 \end{bmatrix}^T$$

For each decision $x$ and each realization of its performance $f^i$ in $\mathcal{F}$, one can compute the utility value $u^j(f^i(x))$ that can be presented in the form of a matrix $\mathbf{U}(x)$ with elements $u^j(f^i(x))$ for $i$ and $j$.

$$\mathbf{U}(x) = \begin{bmatrix} u^1(f^1(x)) & u^2(f^1(x)) & u^3(f^1(x)) & u^4(f^1(x)) \\ u^1(f^2(x)) & u^2(f^2(x)) & u^3(f^2(x)) & u^4(f^2(x)) \\ u^1(f^3(x)) & u^2(f^3(x)) & u^3(f^3(x)) & u^4(f^3(x)) \end{bmatrix}$$

Assuming that $x = (0.5, 0.7)$, one can compute the entries of matrix $\mathbf{U}(x)$, getting:

$$\mathbf{U}(0.5, 0.7) = \begin{bmatrix} 0.6400 & 0.6000 & 0.5400 & 0.5200 \\ 0.8337 & 0.7975 & 0.7433 & 0.7252 \\ 0.3151 & 0.2965 & 0.2686 & 0.2593 \end{bmatrix}$$

In order to compute the expected utility value $E(u(f(x))$ of decision $x$, we first need to compute the matrix:

$$\mathbf{V}(x) = \mathbf{U}(x) \times \Pi = [\big(u^j(f^i(x) \cdot \pi_{i,j}\big)_{\substack{i=1,\ldots,k \\ j=1,\ldots,\ell}}].$$

In our example, we get:

$$\mathbf{V}(0.5, 0.7) = \begin{bmatrix} 0.1280 & 0.0900 & 0.0540 & 0.0260 \\ 0.0667 & 0.0479 & 0.0297 & 0.0145 \\ 0.0378 & 0.0267 & 0.0161 & 0.0078 \end{bmatrix}$$

Then, the expected utility value $E(u(f(x)))$ is obtained as:

$$E(u(f(x))) = \sum_{i=1}^{k} \sum_{j=1}^{\ell} u^j(f^i(x)) \cdot \pi_{ij}. \tag{1}$$

In our example, for $x = (0.5, 0.7)$, the expected utility value is $E(u(f(0.5, 0.7))) = 0.5452$. The variance is given by:

$$\sigma^2(u(f(x))) = \sum_{i=1}^{k} \sum_{j=1}^{\ell} (u^j(f^i(x)) - E(u(f(x))))^2 \cdot \pi_{ij}, \tag{2}$$

which, in our example, gives $\sigma^2(u(f(x))) = 0.0339$.

In general, the DM will try to maximize the expected value $E(u(f(x)))$ and to minimize the variance of the selected solution $\sigma^2(u(f(x)))$. This principle can be applied in different procedures to select a solution $x$ from a set of feasible solutions $X \in R^n$, such as:

- select a solution $x \in X$ with the maximum expected utility value $E(u(f(x)))$ provided that its variance $\sigma^2(u(f(x)))$ is not greater than a given threshold $\sigma^2*$;
- select a solution $x \in X$ with the minimum variance $\sigma^2(u(f(x)))$ provided that its expected utility value is not smaller than a given threshold $E*$;
- select a solution $x \in X$ maximizing a scoring function $S(E(u(f(x))), \sigma^2(u(f(x))))$ being not decreasing with respect to the expected utility value $E(u(f(x)))$ and not increasing with respect to the variance $\sigma^2(u(f(x)))$, as it is the case of

$$S(E(u(f(x))), \sigma^2(u(f(x)))) = E(u(f(x))) - \lambda \cdot \sigma^2(u(f(x)))$$

where $\lambda \geq 0$ is a coefficient representing a DM's aversion to risk.

Let us apply the above procedures to a set of feasible solutions $X = \{x^1, x^2, x^3, x^4\}$, where

- $x^1 = (0.5, 0.7)$,
- $x^2 = (0.8, 0.4)$,
- $x^3 = (0.4, 0.8)$,
- $x^4 = (0.9, 0.2)$.

Let us observe that solution $x^1$ is the same as solution $x$ considered in the above simple example. Computing the expected utility value and the variance for each solution from $X$ we get

- $E(u(f(x^1))) = 0.5452,\ \sigma^2(u(f(x^1))) = 0.0339$,
- $E(u(f(x^2))) = 0.5768,\ \sigma^2(u(f(x^2))) = 0.0350$,
- $E(u(f(x^3))) = 0.5496,\ \sigma^2(u(f(x^3))) = 0.0323$,
- $E(u(f(x^4))) = 0.5643,\ \sigma^2(u(f(x^4))) = 0.0377$.

Consequently:

- if the DM wants to select a solution $x \in X$ with the maximum expected utility value $E(u(f(x)))$ provided that its variance $\sigma^2(u(f(x)))$ is not greater than the threshold $(\sigma^*)^2 = 0.0340$, then solution $x^3$ is selected;
- if the DM wants to select a solution $x \in X$ with the minimum variance $\sigma^2(u(f(x)))$ provided that its expected utility value is not smaller than the threshold $E^* = 0.55$, then solution $x^2$ is selected;
- if the DM wants to select a solution $x \in X$ maximizing a scoring function

$$S(E(u(f(x))), \sigma^2(u(f(x)))) = E(u(f(x))) - 2 \cdot \sigma^2(u(f(x))),$$

then we get

  - $S(E(u(f(x^1))), \sigma^2(u(f(x^1)))) = 0.4773$,
  - $S(E(u(f(x^2))), \sigma^2(u(f(x^2)))) = 0.5068$,
  - $S(E(u(f(x^3))), \sigma^2(u(f(x^3)))) = 0.4850$,
  - $S(E(u(f(x^4))), \sigma^2(u(f(x^4)))) = 0.4890$,

  so that solution $x^2$ is selected.

Another problem that can be considered in this context is the following. Suppose the DM wants to select one solution from $X \subseteq R^n$, which would maximize the expected utility value $E(u(f(x)))$ and minimize the variance $\sigma^2(u(f(x)))$, taking into account a number of

**Table 2** A representation of Pareto-optimal solutions.

| $x_1$ | $x_2$ | Expected value | Variance |
|---|---|---|---|
| 0.388 | 0.862 | 0.580 | 0.030 |
| 0.351 | 0.899 | 0.585 | 0.031 |
| 0.314 | 0.936 | 0.592 | 0.031 |
| 0.302 | 0.948 | 0.594 | 0.032 |
| 0.292 | 0.958 | 0.597 | 0.032 |
| 0.284 | 0.966 | 0.598 | 0.033 |
| 0.276 | 0.974 | 0.600 | 0.033 |
| 0.270 | 0.980 | 0.602 | 0.034 |
| 0.263 | 0.987 | 0.603 | 0.035 |
| 0.258 | 0.992 | 0.605 | 0.035 |
| 0.252 | 0.998 | 0.606 | 0.036 |
| 0.250 | 1.000 | 0.607 | 0.036 |

constraints concerning decision variables $h_s(x) \leq 0,\ s = 1, \ldots, S$. Formally, this problem can be formulated as follows:

maximize: $E(u(f(x)))$

minimize: $\sigma^2(u(f(x)))$

subject to the constraints

$$x \in X, \tag{3}$$

$$h_s(x) \leq 0,\ s = 1, \ldots, S. \tag{4}$$

Obviously, in general, it is not possible to get an optimum value of $E(u(f(x)))$ and $\sigma^2(u(f(x)))$ for the same feasible $x$. Instead, one gets a set of Pareto-optimal solutions $x$, i.e., all solutions $x \in X$ satisfying $h_s(x) \leq 0,\ s = 1, \ldots, S$, for which there does not exist any other solution $\overline{x} \in X$ satisfying $h_s(\overline{x}) \leq 0,\ s = 1, \ldots, S$, having not worse expected utility value $E(u(f(\overline{x})))$ and not worse variance $\sigma^2(u(f(\overline{x})))$, with at least one of the two being better, that is

$$E(u(f(\overline{x}))) > E(u(f(x))), \tag{5}$$

$$\sigma^2(u(f(\overline{x}))) \leq \sigma^2(u(f(x))) \tag{6}$$

or

$$E(u(f(\overline{x}))) \geq E(u(f(x))), \tag{7}$$

$$\sigma^2(u(f(\overline{x}))) < \sigma^2(u(f(x))). \tag{8}$$

Coming back to our example, we have $X = [0, 1]^2$, and let us consider the constraint $h(x) = x_1 + x_2 - 1.25 \leq 0$. Taking into account the set of objective functions $\mathcal{F}$ and the set of utility function $\mathcal{U}$ with respective probability distributions $p$ and $q$, generating the conjoint probability distribution $\Pi$ on $\mathcal{F} \times \mathcal{U}$ introduced above, we can get a set of representative Pareto-optimal solutions presented in Table 2.

**Expected utility value and variance of a set of solutions.** Suppose we have a set of solutions $X = \{x^1, \ldots, x^r, \ldots, x^t\} \subseteq R^n$. In this case, it is possible to compute the expected utility value and the variance of this population of solutions, as follows:

$$E(u(f(X)) = \sum_{r=1}^{t} \sum_{i=1}^{k} \sum_{j=1}^{\ell} u^j(f^i(x^r)) \cdot \pi_{ij} \tag{9}$$

$$\sigma^2(u(f(X))) = \sum_{r=1}^{t} \sum_{i=1}^{k} \sum_{j=1}^{\ell} (u_j(f^i(x^r)) - E(u(f(X))))^2 \cdot \pi_{ij} \tag{10}$$

The expected utility value $E(u(f(X)))$ and the variance $\sigma^2(u(f(X)))$ can be computed using expected utility values and variances of particular solutions in the population, as well as covariances between these solutions:

$$E(u(f(X))) = \sum_{r=1}^{t} E(u(f(x^r))) \tag{11}$$

$$\sigma^2(u(f(X))) = \sum_{r=1}^{t} \sigma^2(u(f(x^r))) + 2 \sum_{r<s} \sigma(u(f(x^r)), u(f(x^s))) \tag{12}$$

where $\sigma(u(f(x^r)), u(f(x^s)))$, $r, s = 1, \ldots, t, r < s$, is the covariance between $u(f(x^r))$ and $u(f(x^s))$, that can be computed as follows:

$$\begin{aligned} \sigma(u(f(x^r)), u(f(x^s))) \quad &= \textstyle\sum_{i=1}^{k} \sum_{j=1}^{\ell} (u_j(f^i(x^r)) - E(u(f(x^r)))) \\ &\quad \cdot (u_j(f^i(x^s)) - E(u(f(x^s)))). \end{aligned} \tag{13}$$

The concepts of the expected utility value and the variance of a set of solution can be applied in multi-objective optimization algorithms with a different aim, for example:

- find a subset of solutions $Y \subset X$ of a given cardinality $q$ having the maximum expected utility value $E(u(f(Y)))$, provided that its variance $\sigma^2(u(f(Y)))$ is not greater than a given threshold $\overline{\sigma}^2$; the subset $Y$ can be found by solving the following $0-1$ quadratic programming problem:

  maximize: $\displaystyle\sum_{r=1}^{t} y_r E(u(f(x^r)))$

  subject to the constraints

  $$\sum_{r=1}^{t} y_r \sigma^2(u(f(x^r))) + 2 \sum_{r=1}^{t-1} \sum_{s=r+1}^{t} y_r y_s \sigma(u(f(x^r)), u(f(x^s))) \leq \overline{\sigma}^2, \tag{14}$$

  $$\sum_{r=1}^{t} y_r = q, \tag{15}$$

  $$y_r \in \{0, 1\}, \; r = 1, \ldots, t; \tag{16}$$

  the optimal subset $Y$ will be composed of $q$ solutions $x^r \in X$ with $y_r = 1$;
- find a subset of solutions $Y \subset X$ of a given cardinality $q$ having the minimum variance $\sigma^2(u(f(Y)))$, provided that the its expected value $E(u(f(Y)))$ is not smaller than a given threshold $\overline{E}$; the subset $Y$ can be found by solving the following $0-1$ quadratic programming problem:

  minimize: $\displaystyle\sum_{r=1}^{t} y_r \sigma^2(u(f(x^r))) + 2 \sum_{r=1}^{t-1} \sum_{s=r+1}^{t} y_r y_s \sigma(u(f(x^r)), u(f(x^s)))$

subject to the constraints

$$\sum_{r=1}^{t} y_r E(u(f(x^r))) \geq \overline{E}, \tag{17}$$

$$\sum_{r=1}^{t} y_r = q, \tag{18}$$

$$y_r \in \{0,1\}, r = 1,\ldots,t; \tag{19}$$

again, the optimal subset $Y$ will be composed of $q$ solutions $x^r \in X$ with $y_r = 1$.

Coming back to our example, let us consider again the solutions from the set $X = \{x^1, x^2, x^3, x^4\}$, and let us compute the covariances $\sigma(u(f(x^r)), u(f(x^s)))$, obtaining the following variance-covariance matrix $\mathbf{\Sigma}(X) = [\sigma(u(f(x^r)), u(f(x^s)))]$, where $\sigma(u(f(x^r)), u(f(x^r))) = \sigma^2(u(f(x^r)))$:

$$\mathbf{\Sigma}(X) = \begin{bmatrix} 0.0339 & 0.0258 & 0.0318 & 0.0157 \\ 0.0258 & 0.0350 & 0.0182 & 0.0334 \\ 0.0318 & 0.0182 & 0.0323 & 0.0064 \\ 0.0157 & 0.0334 & 0.0064 & 0.0377 \end{bmatrix}$$

Let us suppose that the DM wants to select a subset of solutions $Y \subset X$ with cardinality $q = 3$, having the maximum expected utility value $E(u(f(Y)))$. Solving the 0-1 quadratic programming problem presented above, and without considering any constraint on the variance $\sigma^2(u(f(Y)))$, we get that the DM has to select the subset $Y_1 = \{x^2, x^3, x^4\}$ with expected utility value $E(u(f(Y_1))) = 1.6907$ and variance $\sigma^2(u(f(Y_1))) = 0.2211$.

If, in turn, the DM would like to select a subset of solutions $Y \subset X$ with cardinality $q = 3$, having the minimum variance $\sigma^2(u(f(Y)))$, then, by solving the corresponding 0-1 quadratic programming problem presented above, and without considering any constraint on the expected value $E(u(f(Y)))$, the DM would get the subset $Y_2 = \{x^1, x^3, x^4\}$ with expected utility value $E(u(f(Y_2))) = 1.6591$ and variance $\sigma^2(u(f(Y_2))) = 0.2118$.

Suppose now that the DM would like to select a subset of solutions $Y \subset X$ with cardinality $q = 2$, having the maximum expected utility value $E(u(f(Y)))$ but under the condition that the variance $\sigma^2(u(f(Y)))$ is not greater than 0.215. In this case, solving the corresponding 0-1 quadratic programming problem, the DM would get the subset $Y_3 = \{x^3, x^4\}$ with expected utility value $E(u(f(Y_3))) = 1.1139$ and variance $\sigma^2(u(f(Y_3))) = 0.0828$.

Finally, suppose that the DM would like to select a subset of solutions $Y \subset X$ with cardinality $q = 2$, having the minimum variance $\sigma^2(u(f(Y)))$ but under the condition that the expected utility value $E(u(f(Y)))$ is not smaller than 1.1. In this case, the DM would get again the subset $Y_4 = \{x^3, x^4\}$.

The above two problems of selecting a subset of solutions of a given cardinality maximizing the expected utility value with a constraint on the variance, or minimizing the variance with a constraint on the expected value, can be interpreted as a discrete version of the Markowitz portfolio selection problem in the context of multi-objective optimization. It is sensible to consider also the classic continuous Markowitz portfolio selection problem which consists in searching for a vector

$$\mathbf{y} = [y_1, \ldots, y_t],\ y_r \geq 0,\ r, \ldots, t,\ \sum_{r=1}^{t} y_r = 1,$$

that maximizes the expected utility value

$$E(u(f(\mathbf{y}))) = \sum_{r=1}^{t} y_r E(u(f(x^r)))$$

subject to the constraint that the variance $\sigma^2(u(f(\mathbf{y})))$ is not greater than a given threshold $\overline{\sigma}^2$, that is

$$\sigma^2(u(f(\mathbf{y}))) = \sum_{r=1}^{t} y_r \sigma^2(u(f(x^r))) + 2\sum_{r=1}^{t-1}\sum_{s=r+1}^{t} y_r y_s \sigma(u(f(x^r)), u(f(x^s))) \leq \overline{\sigma}^2.$$

The classic Markowitz portfolio selection problem can also be formulated as minimization of the variance $\sigma^2(u(f(\mathbf{y})))$ under the constraint that the expected utility value $E(u(f(\mathbf{y})))$ is not smaller than a given threshold $\overline{E}$.

Coming back to our example, let us suppose that the DM wants to compute the vector $\mathbf{y} = [y_1, \ldots, y_4]$ having the maximum expected utility value $E(u(f(\mathbf{y})))$ but under the condition that the variance $\sigma^2(u(f(\mathbf{y})))$ is not greater than 0.o25. In this case, the optimal vector is

$$\mathbf{y^1} = [0 \quad 0.4223 \quad 0.3487 \quad 0.2289],$$

with its corresponding expected utility value $E(u(f(\mathbf{y^1}))) = 0.5644$ and variance $\sigma^2(u(f(\mathbf{y}^1))) = 0.025$.

Instead, if we suppose that the DM wants to compute a vector $\mathbf{y} = [y_1, \ldots, y_4]$ having the minimum variance $\sigma^2(u(f(\mathbf{y})))$ but under the condition that the expected utility value $E(u(f(\mathbf{y})))$ is not smaller than 0.56, then the optimal vector is

$$\mathbf{y^2} = [0 \quad 0.1599 \quad 0.4285 \quad 0.2118]$$

with its corresponding expected utility value $E(u(f(\mathbf{y}^2))) = 0.56$ and variance $\sigma^2(u(f(\mathbf{y}^2))) = 0.0224$.

Let us finally remark, that the value of $y_r, r = 1, \ldots, t$, can be interpreted as a score assigned by a fitness function to the corresponding solution $x^r$ in an evolutionary optimization algorithm, such that the greater the value of $y_r$ the more probably $x^r$ should be selected to generate a new solution.

**Heat map visualization of averages and variances.** For a visualization of the situation that is described above consider Figure 2. For any two-dimensional input/decision/design/output variable $x = (x_1, x_2)$ in the domain $[0, 1] \times [0, 1]$, we can compute the mean and variance of the $l \cdot k$ (here, $3 \cdot 4 = 12$) entries of the resulting matrix $U(x)$ or $U(y)$. Then, the figure on its left and right side shows the thus computed mean values and variances for variables $x$ or $y$ of a discretized grid on $[0, 1] \times [0, 1]$.

### 4.1.4 Application to sea-level rise and storm surge projections

This section describes a real-world application regarding the deep uncertainties in sea-level rise and storm surge projections. This example represents a probabilistic generalization of the classical Van Dantzig decision analytical application where the decision is to choose the level of increase in dike height to reduce flood risk [1]. The two objectives are probabilistic as a function of uncertainties in sea level rise due to climate change and local effects of the geophysics of storm surge (i.e., two different but interdependent geophysical models). Figure 3 illustrates the original deterministic Van Dantzig baseline, the mean trade-off between flood risk and investment, as well as the relative locations of the minimum net present values for investment. The challenge as emphasized in the log scale zoomed view is the mean Pareto front would not provide a DM an understanding of the severe variance

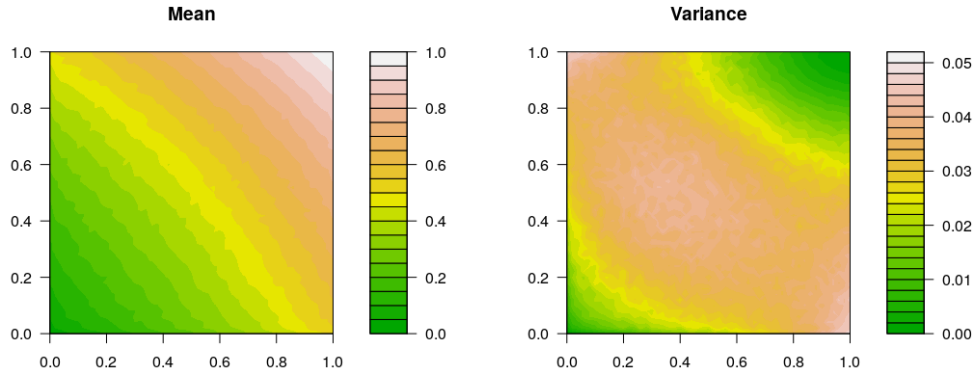



**Figure 2** Heat map visualization of averages (on the left) and variances (on the right).

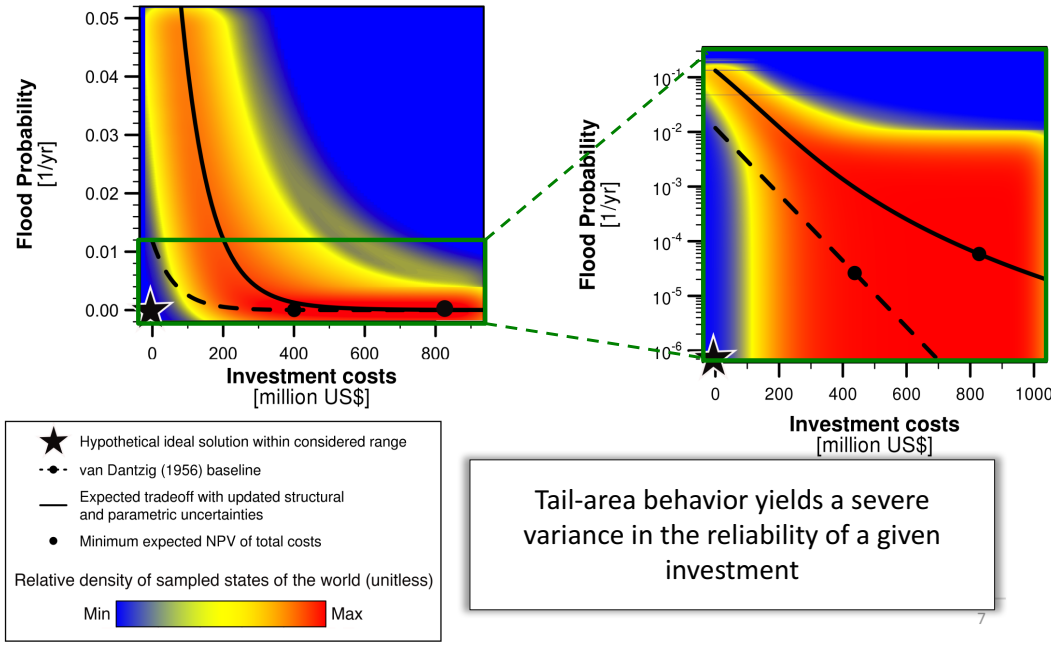



**Figure 3** Real-world application about uncertainty in technical information (adapted from [1]).

in the potential outcomes for a given investment. For example, working with the mean trade-off an investment 800 Million US Dollars intended to provide a 1 in 10,000 year level of flood protection has a significant residual probability of dramatically less protection (severe damages and potential loss of life). This probabilistic Pareto space context poses a challenge to decision making, particularly given the potential uncertainties in preferences or risk aversion for the residual risks. It then motivates the question of understanding the potential joint probabilistic outcome of uncertain Pareto performance and uncertain DM's preferences.

### 4.1.5 Open questions

In this report, we proposed a novel approach for interactive multi-objective optimization taking into account uncertainty referring to both the evaluations of solutions by objective functions as well as the preferences of the decision maker. We envisage the following directions for future research.

Firstly, we aim at developing methods for elicitation of probability distributions on objective performances and on utility functions. Secondly, we will propose some procedures for robustness analysis that would quantify the stability of results (utilities, ranks, and pairwise relations) obtained in view of uncertain performances and preferences. Thirdly, when aiming to select a set of feasible options, we will account for the interactions between different solutions. Fourthly, we will integrate the proposed methods with evolutionary multi-objective optimization algorithms with the aim of evaluating and selecting a population of solutions. Fifthly, we plan to adapt the introduced approach to a group decision setting, possibly differentiating between two groups of decision makers being responsible for, respectively, setting the goals and compromising these goals based on different utilities. Finally, we will apply the proposed methodology to real-world problems with highly uncertain information about the solutions' performances and decision makers' preferences.

#### References

1 P. C. Oddo, B. S. Lee, G. G. Garner, V. Srikrishnan, P. M. Reed, C. E. Forest, and K. Keller. Deep uncertainties in sea-level rise and storm surge projections: Implications for coastal flood risk management. *Risk Analysis*, page (in press).

2 B. Roy. Main sources of inaccurate determination, uncertainty and imprecision in decision models. *Math. Comput. Modelling*, 12:1245–54, 1989.

3 B. Roy. Do not confuse. *Int. J. Multicriteria Decision Making*, 6:112–17, 2016.

4 I. Yevseyeva, A. P. Guerreiro, T. T. M. Emmerich, and C. M. Fonseca. A portfolio optimization approach to selection in multiobjective evolutionary algorithms. In *International Conference on Parallel Problem Solving from Nature*, pages 672–681. Springer, 2014.

## 4.2 Personalization of multicriteria decision support systems (WG2)

*Matthias Ehrgott, Gabriele Eichfelder, Karl-Heinz Küfer, Christoph Lofi, Kaisa Miettinen, Luís Paquete, Stefan Ruzika, Serpil Sayın, Ralph E. Steuer, Theodor J. Stewart, Michael Stiglmayr, and Daniel Vanderpooten*



**Abstract.** In this report, personalization is approached from a learning perspective. We propose a framework for a decision support system to help a decision maker who faces the problem of identifying a most preferred from among a set of alternatives. Our framework encompasses the idea that the objectives and the constraints of the model may not be clear at the beginning and are likely to evolve throughout the decision process. Our proposal deviates from the vast literature on interactive methods by allowing the model to evolve in a very flexible way. We illustrate the need of personalized decision support systems with some applications. We also discuss ways to present solutions to a decision maker in a qualitative manner as this is an important part of the iterative learning and solution process.

### 4.2.1 Introduction

We approach personalization from a learning perspective and propose a framework for a decision support system to help a decision maker (DM) who faces the problem of identifying a most preferred solution from among a set of alternatives. Our framework is general in the sense that it allows for a continuous and discrete expression of alternatives. The alternatives may be explicitly available or may be defined implicitly via some functions (objectives and constraints). Our framework encompasses the idea that the objectives and the constraints of the model may not be clear at the beginning and are likely to evolve throughout the decision process. Thus, the process by which the DM modifies his/her perception of preferences through restructuring of the hierarchical decision model must be facilitated. This can be achieved, for instance, by adding/subtracting objectives, aggregating/disaggegating objectives, modifying constraints, converting constraints into objectives and vice versa while retaining insights gained from earlier phases of the analysis. Figure 4 illustrates the iterative decision making process. Some Pareto optimal solutions of an initial model are studied by a DM. These solutions reveal some findings about the problem to the DM or help him/her discover one's preferences. These are taken into account in a revised model and some carefully revised new Pareto optimal solutions are presented to the DM on the next round, and so forth. The process continues until the DM identifies a most preferred solution.

Our proposal deviates from the vast literature on interactive methods by allowing the model to evolve in its degree of flexibility. As the objectives and constraints of the model are modified, the Pareto optimal set shifts and changes. We have seen studies in the literature in which the solution method is switched depending on the phase of the solution process, i.e., the search. However, in these studies the model usually stays the same. Here, we understand personalization as enabling the model to evolve.

The influence of adding and subtracting objective functions to a multiobjective optimization problem has been considered in [9]. Furthermore, the relative importance of objectives is discussed and a definition of weights is given in [19, 20] (where weights are called coefficients) for objective functions as well as for groups of objective functions. This approach results in a convex combination of functions similar to linear combinations as discussed in [3]. There, strategies are discussed that reduce the size of the solution set of the multiobjective optimization problem for instance by combining several objectives linearly, i.e. by summing them up, before employing tools to solve the resulting multiobjective optimization problem. Using partial preference models, where weights are partially defined, is also a way of focusing on reduced solution sets of interest [13].

The need of iterating to find an appropriate model of a real-world problem to be solved is demonstrated in [2, 26] with cases in optimal shape design of an air intake channel and a two-stage separation process, respectively. In the latter case, an interactive multiobjective optimization method helped in validating and improving the model and only after that kind of iterating, the actual interactive solution process was conducted.

The idea of constraint optimization using multiobjective optimization models, i.e. the idea to transform constraints to objectives, as well as the other way around, is studied in [15]. Furthermore, e.g., in [16], it is demonstrated that converting a problem with one objective and four very demanding constraints can be solved by optimizing constraint violations besides the original objective, i.e., a problem with five objectives. Hence, in the literature, the relation between constrained and multiple objectives as well as between aggregated and disaggregated multiobjective optimization problems is already studied at least in parts, while several such models have so far not been used in an iterative manner on varying levels for steering a decision-making process.

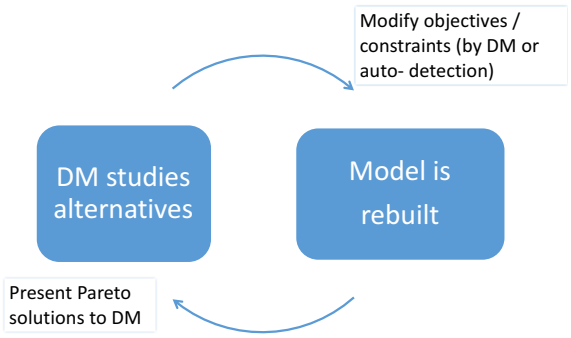



**Figure 4** The framework for personalized decision support.

In [12], an unconstrained bi-objective discrete optimization problem is studied with the goal of finding representations that adhere to a given quality with respect to the $\epsilon$-indicator measure. The suggested approach is related to the Nemhauser-Ullman algorithm that has been proposed for the traditional knapsack problem which has one objective function and one constraint. The work of [23] brings this idea closer to the discussion in this report by formulating a bi-dimensional knapsack problem where one of the constraints is a soft constraint. The authors model the soft constraint as an objective function, thereby ending up with a uni-dimensional knapsack problem with two objectives. As such, they propose to compute representative solutions for the transformed problem so as to portray the trade-off between the objective function of the original problem and satisfaction or violation of its soft constraint.

In [11, Section 2] and [5, Section 3], a detailed review on the literature on modeling the relative importance of objectives is provided. In these references, as well as in [6], partial orderings, other than the natural orderings via (non)polyhedral cones, are examined for their impact on optimal (in that case, efficient) solution sets of multiobjective optimization problems. These examinations might help in understanding the relationship between (dis)aggregated multiobjective optimization problems.

## Problem formulation

To give a mathematical formulation of the problem of adding/subtracting and (dis)aggregating objectives, we make the following assumptions:

- Let a nonempty subset $X \subseteq \mathbb{R}^n$ be given which describes the set of alternatives. For instance, the set $X$ might be determined by some hard constraints given by laws of nature, which cannot be weakened and, thus, cannot be transformed to objective functions.
- Let $\mathcal{F} := \{f_i \colon \mathbb{R}^n \to \mathbb{R} \mid i = 1, \ldots, k\}$ be a finite collection of functions which are potentially of interest for particular models. Then, for particular model instances, some of these functions can appear in the formulation of the objective functions or in the constraints.
- Let $h_1, \ldots h_m,\ g_1, \ldots, g_l \colon \mathbb{R}^k \to \mathbb{R}$ be arbitrary functions describing which of the functions $f \in \mathcal{F}$ are aggregated or chosen for the formulation of the individual objective functions or constraints of the particular model. Thereby, $m \in \mathbb{N}$ and $l \in \mathbb{N}$ also depend on the particular model instance.

Under these assumptions, a particular model instance can be expressed as

$$\min_{x\in S} h_1(f_1(x),\dots,f_k(x)),\dots,h_m(f_1(x),\dots,f_k(x)), \tag{PMI}$$

where $S := \{x \in X \mid g_i(f_1(x),\dots,f_k(x)) \leq 0,\ i = 1,\dots,l\}$.

▶ **Example 1.** Let $X = \mathbb{R}^n$ and $\mathcal{F} = \{f_1, f_2, f_3 \colon \mathbb{R}^n \to \mathbb{R}\}$. For the aggregation functions $h_j$ we only take linear combinations and selections into account. Thus, let weights $w_2, w_3 > 0$ be given. With $h_1(y_1, y_2, y_3) = y_1$ and $h_2(y_1, y_2, y_3) = w_2y_2 + w_3y_3$ we get

$$\min_{x\in X} \begin{pmatrix} f_1(x) \\ w_2f_2(x) + w_3f_3(x) \end{pmatrix}. \tag{$P_A(w_2, w_3)$}$$

The corresponding disaggregated multiobjective optimization problem with functions $h_1(y_1, y_2, y_3) = y_1$, $h_2(y_1, y_2, y_3) = y_2$, and $h_3(y_1, y_2, y_3) = y_3$ is

$$\min_{x\in X} \begin{pmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{pmatrix}. \tag{$P_D$}$$

When disaggregating the problem $(P_A(w_2, w_3))$ one might be interested in keeping the properties of an already found Pareto optimal solution $\bar{x} \in X$ of the bi-objective problem $(P_A(w_2, w_3))$. For instance, it might be the aim to keep the achieved level for the value $f_1(\bar{x})$ while being willing to explore nearby values for $f_2$ and $f_3$. With $g_j(y_1, y_2, y_3) = y_j - \Delta_j$ for $j = 1, 2, 3$ and with

$$\Delta_1 = f_1(\bar{x}), \qquad \Delta_2 = \Delta_3 = w_2f_2(\bar{x}) + w_3f_3(\bar{x}) + \delta$$

for some scalar $\delta \geq 0$, also the following problem might be of interest.

$$\begin{aligned} \min\ & \begin{pmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{pmatrix} \\ & \text{s.t.} \\ & f_1(x) \leq \Delta_1, \\ & f_2(x) \leq \Delta_2, \\ & f_3(x) \leq \Delta_3, \\ & x \in X, \end{aligned} \tag{$P_C(\Delta)$}$$

where we can write $S = \{x \in X \mid f_i(x) \leq \Delta_i,\ i = 1, 2, 3\}$.

The following relations are, for instance, obvious:

- If a point $\bar{x} \in X$ is Pareto optimal for $(P_D)$, then $\bar{x}$ is also Pareto optimal for $(P_C(\Delta))$ for any $\Delta \in \mathbb{R}^3$ with $\Delta_i \geq f_i(\bar{x})$, $i = 1, 2, 3$.
- If a point $\bar{x} \in X$ is Pareto optimal for $(P_C(\Delta))$ for any $\Delta \in \mathbb{R}^3$, then $\bar{x}$ is also Pareto optimal for $(P_D)$.
- If a point $\bar{x} \in X$ is Pareto optimal for $(P_A(w_2, w_3))$ for some weights $w_2, w_3 > 0$, then $\bar{x}$ is also Pareto optimal for $(P_D)$.

Interesting questions are also, for instance, under which assumptions a Pareto optimal point $\bar{x}$ of $(P_A(w_2, w_3))$ for some weights $w_2, w_3 > 0$ is at least feasible for $(P_C(\Delta))$ for $\Delta_1 \geq f_1(\bar{x})$, $\delta \geq 0$ and

$$\Delta_2 = \Delta_3 = w_2f_2(\bar{x}) + w_3f_3(\bar{x}) + \delta\ .$$

### 4.2.2 Applications

Next we illustrate the need of personalized decision support systems with some applications.

#### Radiotherapy

In radiotherapy, the set of alternatives consists of applicable treatment plans $x$. We assume that the alternatives are judged by the DM solely based on the properties of the resulting dose distribution. At the highest level, the properties of the dose distribution predict the likelihood of treatment success or failure as well as the likelihood of specific complications and side effects related to the organs at risk. To represent and compute the dose distribution, the patient image is divided into (up to millions of) equal-sized voxels. The dose distribution $D(x)$ is then the vector of all voxel dose values. For the sake of simplicity, each voxel either belongs to a target, to a specific organ at risk, or to normal tissue. For each target, there is a prescribed dose $d^{presc}$ that is deemed adequate to kill all tumor cells. For evaluating a given dose distribution, a large collection of objective functions has been established in the radiotherapy community. Most of these objective functions in some way measure the average under- or overdose over all voxels belonging to a specific structure (target or organ at risk). However, other ("lower-level") aspects of the dose distribution also play a role, such as smallish localized areas of too high dose far away from the target ("hot spots"). This is where aggregation and disaggregation come into play.

**Aggregation and disaggregation in radiotherapy.** The dose values in the individual voxels form a natural basis of lowest level and highest detail when assessing the dose distribution. The following implications can be assumed to hold for any DM's utility function:

- For target voxels $i$, as long as the dose values are below the prescribed dose, $d_i(x) < d_i(x')$ and all else equal, this implies that $x$ is a worse treatment plan than $x'$.
- For target voxels $i$, as long as the dose values are above the prescribed dose, $d_i(x) < d_i(x')$ and all else equal, this implies that $x$ is a better treatment plan than $x'$.
- For risk and normal tissue voxels $j$, $d_j(x) < d_j(x')$ and all else equal, this implies that $x$ is a better treatment plan than $x'$.

Fundamental ("atomic" or "lowest-level") objective functions $\mathcal{F}$ can be chosen as representations of these relations:

- For target voxels $i$: $f_i^{UD}(x) = \max\{0, d^{presc} - d_i(x)\}$.
- For target voxels $i$: $f_i^{OD}(x) = \max\{0, d_i(x) - d^{presc}\}$.
- For risk and normal tissue voxels $j$: $f_j(x) = d_j(x)$.

A decision process based on $\mathcal{F}$ is infeasible. Given two unrelated dose distributions, a comparison may well exceed the mental capacity of a DM. Even if a trajectory is provided where in each comparison only a few fundamental functions differ, the search space would be too large and any search too unstructured for efficient decision making. Thus, "higher-level" functions are introduced that aggregate all fundamental functions of voxels of the same structure, for example, the squared organ at risk dose:

$$f_{risk}(x) = \sum_{j \in risk} (d_j(x))^2. \tag{20}$$

The aggregation simplifies the problem by treating every voxel within the structure as equal, disregarding position and spatial relationship to other voxels. Also, it handles the trade-off within the voxels of the same structure automatically, depending on the exact formulation of the aggregation (which can be chosen by the DM).

On the other hand, the aggregated function can cloud lower-level aspects of the DM's utility function. For example, the DM may be happy with the overall amount of dose for the organ at risk, but there is a certain region inside the organ at risk that still gets too much dose. One option would be to choose a different aggregated function, maybe using a higher coefficient in order to penalize higher doses more and force a different trade-off of voxel doses inside the organ at risk.

However, the discontent may be attributed more to the specific location and the spatial accumulation of higher dosed voxels, rather than the values themselves. In this case, the assumptions made when aggregating the fundamental functions – namely that all voxels are equally independent of location and spatial relationship to other voxels – breaks down. In this case, lower-level functions may need to be (re-)introduced in the variable model, i.e. the model must be disaggregated.

### Land use planning

Land use planning involves the allocation of facilities to specific locations or activities to specific areas within a region of land. In most non-trivial contexts, land-use planning involves many criteria, some at least of which will involve partially qualitative considerations such as social impacts of displacements, destruction of old burial sites and effects of biodiversity reduction. Typically also, conflict is generated between multiple stakeholders that needs some resolution before any decision can be implemented.

Two examples of land use planning problems with which one of the authors has been associated are the following. The first related to replacement of indigenous afromontane grasslands on the eastern escarpment areas of South Africa by exotic commercial forestries [27]. The prime decision variables related to proportions of the region allocated to forestry, with subsidiary considerations including water supply to rural communities for subsistence and agriculture, and preservation of biodiversity in the region. The second example arose from restoration of land for nature conservation with associated partitioning of land into intensive and extensive agriculture, as well as other development activities, in the Netherlands [4]. The prime decision variables were binary, i.e. selection of activity for each designated parcel of land.

Land use planning provides a challenging context within which to seek personalization of decision support. Different stakeholders will have different perspectives on the same problem, which need to be provided for. As different groups work together and negotiate, problem structures and preference perceptions evolve dynamically, and this too needs to be captured in the decision support system.

Some dynamic issues which arose in these examples included the following:

- A need to incorporate policy (not entirely hard) constraints into the forestry development problem, that for any chosen proportion of area to forestry, the precise locations of the plantations were to be subject to environmental impact vetoes;
- The original decision support models for selection of land parcel activities focused on assessing the value of allocating each activity to each parcel as primary objectives. But deeper reflection led to a realization that system management requires the definition of further system-related criteria concerned with coherency of activities which are non-additively related to decision variables.

- In the water resources component of the South African forestry land allocation, one initially identified criterion was interests of rural village communities. But problems encountered while attempting to evaluate decision alternatives according to this criterion led to a realization that there were two relevant sub-criteria, that could be seen as "female" (close access to clean water) and "male" (availability of piecework on commercial farms).

Any decision support system must be able to cope with such often unexpected developments in the problem structure as regards both the decision space and the set of criteria.

### 4.2.3 Research questions

In the following, we discuss some of the main research questions that need to be addressed in a personalized iterative decision making process as described in previous sections.

#### Aggregating/disaggregating functions as objectives and constraints

We start again by motivating our research questions with an example. Let us consider a problem where a DM wants to minimize cost $f(x)$ and maximize quality $g(x)$ of a product to be purchased:

$$\min f(x), \max g(x).$$

The quality may consist of two separate components: $g(x) = w_1 g_1(x) + w_2 g_2(x)$, cf. Example 1.

Let us suppose that a solution $\bar{x}$ is identified by the DM after a first depiction of the Pareto front (in the objective space) of this problem. Now, the question is to find new solutions, not too far away from $\bar{x}$, of a possible disaggregated problem. Then one might solve the problem

$$\min f(x), \max g_1(x), \max g_2(x)$$

or

$$\begin{aligned} & \min f(x) \\ & \text{s.t.} \\ g_1(x) &\leq g(\bar{x}) + \Delta_1, \\ g_2(x) &\leq g(\bar{x}) + \Delta_2. \end{aligned}$$

Open research questions include:

- Are the relationships between reformulations of the problem stronger if $g_1$ and $g_2$ are somehow correlated? Does the strength of the relationship depend on $\bar{x}$? From a practical point of view, is the non-correlated case of even more interest?
- How can a recommendation for an initial aggregation be made in order to start the decision-making process? How can objectives be added or removed? There can be settings when the model is blank (unknown) or very well-known. In the first case, the model is to be built by adding, in the other, by removing.
- An expressed constraint may be found to be irrelevant after learning that the range is too narrow to be relevant. The question is how to model this automatically.
- An objective can be converted into a constraint to eliminate unwanted alternatives or to save levels with specific objectives. The question is how to structure such approaches and what are the relations between the solutions found.

### Navigation

To form a good base for the selection step of a solution $\bar{x}$ in an iterative process, a good presentation and a way to navigate between possible solutions is required. We state some known approaches as well as some open questions in the following.

**Navigation in a continuous space of alternatives.** For a continuous multiobjective optimization problem, a real-time navigation capability for the DM such as the following two-step process can be offered:

1. Optimizing a set of representative solutions $x_1, \ldots, x_m$ in an offline pre-computation, with objective function vectors $F_i = F(x_i)$. Explicitly or implicitly, the representative pairs $(x_i, F_i)$ must have a neighborhood relationship defined, allowing neighboring solutions to be linearly interpolated. This means that for a subset $I$ of mutually neighboring points, and for coefficients $\lambda_i \geq 0$ with $\sum_{i \in I} \lambda_i = 1$
   - any interpolated solution $x = \sum_{i \in I} \lambda_i x_i$ is feasible,
   - for any interpolated point $x$, the objective function values $F(x)$ differ from the Pareto optimal achievable values only by an acceptable error ("approximation quality"),
   - $F(\sum_{i \in I} \lambda_i x_i) \approx \sum_{i \in I} \lambda_i F_i$ in order for the navigation mechanisms of the item above to work ("triangulation of Pareto front approximation").
2. Searching the space of interpolated solutions in real-time. This can be done by solving linear optimization problems in the interpolation coefficients.

For convex problems, this is understood (see, "sandwiching" [24] for the calculation of the representative solutions, and real-time navigation in [7, 17, 18]), but maybe not published well enough yet. In the convex case, many of the ingredients mentioned above come for free (neighborhood from calculating the convex hull, feasibility of interpolated solutions) or coincide (second and third bullet points as a consequence of sandwiching). For general nonconvex problems, this is not the case. One way of connecting objective and decision spaces for nonconvex problems has been proposed in [10]. Research questions include:

- Formalizing the approach, maybe embedding the convex case as a special case, in order to make it more known and understood in the community.
- Properties of nonconvex problems to facilitate this approach.
- Development, improvement, and description of algorithms for the calculation of representative solutions and for real-time navigation especially for the nonconvex case.

**Navigation in a discrete space of alternatives.** In a discrete case, the DM wants to find the preferred solution out of a finite but typically large set of alternatives. Such a decision problem can also be handled by real-time navigation mechanisms. However, interpolation is not possible. Thus, when traversing a set of alternatives, the direction and size of each navigational step cannot be controlled very well. Research questions include:

- How can the wishes of a DM be stated and interpreted in the context of discrete navigation?
- Should the DM follow a trajectory by jumping from alternative to alternative? If yes, how should the next alternative be chosen? Can this choice be defined by a particular distance measure or neighborhood relationship in the space of alternatives?
- Or should the navigation mechanism focus more on eliminating alternatives?

### 4.2.4 Toward personalizing representations

Personalization is very much related to learning. In different domains, there may be different aspects of learning. Expert DMs may have a good understanding of the structure of a decision problem but they may still need to learn about the nature of the problem instance

(e.g. in radiotherapy) and gain insight in the conflicting nature of the objectives and feasible solutions as well as the feasibility of their preferences. Novice DMs may need to discover their objectives, constraints and solutions.

Throughout this section, we assume that some model (as a result of processes described in the sections before) is given together with an explicit list of $n$ alternatives (e.g. items/products). The properties of these alternatives are described by criteria (defined on measurable scales). This is for example the case in online sales or consulting systems where customers are supported in choosing some product meeting their individual demands.

In many such practical applications, the set of Pareto optimal solutions exceeds a manageable cardinality. In order to analyze or visualize the set of alternatives and, thus, to assist the process of making a final decision, the DM requires a concise representation of the Pareto optimal set to obtain a quick overview. A good representation can still communicate the nature of the set while hiding options which are not informative. In the following, we investigate the influence of personalization on representations, adaptation of quality measures incorporating personal preferences and algorithms to compute a personalized representation in the context of explicitly given alternatives.

An idea to incorporate personalization in the computation of a representative subset is based on two functionalities which can be in principle applied in an arbitrary order during a decision making process:

1. The computation of a good and concise representation for a given region of interest.
2. The determination of the set (or a representation) of neighbors wrt. to a selected point.

During the search for a finally preferred solution, a DM may iteratively make use of these two functionalities: A good representation for the problem/model at hand may be computed and analyzed, the model may be changed and the first functionality may be invoked again, or, eventually, a DM may be interested in the neighborhood of some selected point to be informed about similar alternatives. Before presenting some specific algorithmic ideas, we discuss these two functionalities in more detail first.

Concerning functionality 1, a crucial point relates to the notion of "goodness" of a representation, i.e. the quality of a representation. Certainly, one goal is to determine a representation $R$ of the set of Pareto optimal points (also known as nondominated points) $Y_N \in \mathbb{R}^p$ which is tractable for the DM and can be efficiently computed. We rely on the classical quality measures for discrete representations suggested and discussed in [8, 22], namely coverage, uniformity and cardinality which can be roughly characterized as follow.

- *Coverage*: any point in $Y_N$ is represented or *covered* by at least one point in $R$.
- *Uniformity*, also called *spacing*: any two points in $R$ are sufficiently *spaced*, avoiding redundancies.
- *Cardinality* refers to the cardinality $|R|$ of the representation $R$. Since each representative point has to be computed with a certain effort, the cardinality should be small.

The concepts coverage and *spacing* can be implemented in a variety of ways. In principle, one can distinguish between a *geometric vision* based on *distances* and a *preference-oriented vision* using some *preference relation*. In a geometric vision, distances between points in $Y_N$ and points in R are used to evaluate coverage. Likewise, uniformity is evaluated by calculating pairwise distances between points in $R$. Alternatively, a preference-oriented vision is based on a preference relation $\preceq$. For two points $y$ and $y'$, one can then say that $y$ *covers* $y'$ if $y \preceq y'$ which implies a notion of coverage. Analogously, $y$ and $y'$ are sufficiently *spaced* if not $(y \preceq y')$ and not $(y' \preceq y)$ which then defines the notion of uniformity/spacing.

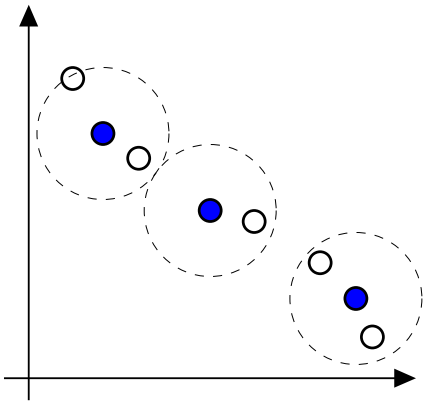


**Figure 5** Illustration of a representation based on coverage.

### 4.2.5 Algorithmic approaches for computing personalized representations

Based on the discussion in the previous section, several methods existing in the literature are proposed, which can be adapted, to meet the two functionalities mentioned. The first three methods are geometric-based approaches, while the fourth one is a preference-based approach. These efforts may be understood as a first attempt of computing personalized representations.

#### A geometric-based approach

In a geometric vision, coverage measures the quality of the representative subset by considering the distance of the unchosen elements to their closest elements in the subset. Formally, the coverage of a subset $R \subseteq Y_N$ is computed as

$$I_C(R, Y_N) = \max_{y \in Y_N} \min_{y' \in R} \|y - y'\|.$$

The coverage representation problem consists of finding a subset of cardinality $k$ that has the smallest coverage value, i.e.,

$$\min_{\substack{R \subseteq Y_N \\ |R|=k}} I_C(R, Y_N).$$

This problem is known as the $k$-center problem [14]. In the particular case of two objectives, it can be solved in a polynomial amount of time [28].

Similarly, in a geometric vision, uniformity measures how far apart the $k$ chosen elements of the set $R \subseteq Y_N$ are from each other. It is computed as the minimum distance between a pair of distinct elements as

$$I_U(R) = \min_{\substack{y, y' \in R \\ y \neq y'}} \|y - y'\|.$$

The goal of the uniformity representation problem is to find a subset $R$, with a given cardinality $k$, from a set $Y_N$ that maximizes $I_U(R)$, i.e.,

$$\max_{\substack{R \subseteq Y_N \\ |R|=k}} I_U(R).$$

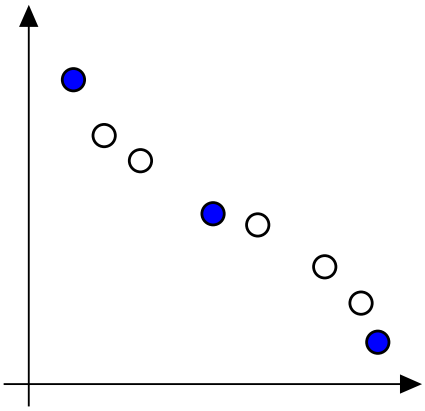


**Figure 6** Illustration of representation based on uniformity.

Note that this problem corresponds to a particular case of the $k$-dispersion problem in facility-location [21]. Also for the particular case of two objectives, this problem can be solvable in a polynomial amount of time [28].

Note, that functionality 2 suggests itself in a geometric vision: The neighborhood for the second functionality is an $\varepsilon'$-neighborhood of a selected point $\bar{y}$:

$$y \in Y_N \ : \ \|y - \bar{y}\| \leq \varepsilon'.$$

### The revised boundary intersection method

The revised boundary intersection (RNBI) method computes a discrete representation of the Pareto optimal set of a multiobjective linear optimization problem (MOLP) $\min\{Cx : Ax \leqq b\}$ with a bounded feasible set. It provides guarantees on both the uniformity and the coverage error of the representation, see [25]. The following is a description of the algorithm.

1. Input: MOLP data $A, b, C$ and $ds > 0$.
2. Find $y^{AI}$ defined by $y_k^{AI} = \max\{y_k : y \in Y\}$ for $k = 1, \ldots, p$.
3. Find a Pareto optimal point $\hat{y}$ by solving the linear problem $\phi := \min\{e^T y : y \in Y\}$.
4. Compute $p+1$ points $v^k, k = 0, \ldots, p$ in $\mathbb{R}^p$

$$v_l^k = \begin{cases} y_l^{AI}, & l \neq k, \\ \phi + \hat{y}_k - e^T v^0 & l = k. \end{cases}$$

   - The convex hull $S$ of $\{v^0, \ldots, v^p\}$ is a simplex containing $Y$.
   - The convex hull $\hat{S}$ of $\{v^1, \ldots, v^p\}$ is a hyperplane with normal $e$ supporting $Y$ in $\hat{y}$.
5. Compute equally spaced reference points $q^i$ with a distance $ds$ on $\hat{S}$.
6. For each reference point $q$ solve the linear problem $\min\{t : q + te \in Y, t \geqq 0\}$ and eliminate dominated points from the resulting set $R$.
7. Output: Representation $R$.

The steps of the algorithm are illustrated in Figure 7.

Theorem 2 provides the quality guarantee for the method in terms of uniformity and coverage error of the generated representatiom.

▶ **Theorem 2.** *Let $R$ be the representation of $Y_N$ obtained with the RNBI method.*

1. *Let $q^1, q^2$ be two reference points with $d(q^1, q^2) = ds$ that yield Pareto optimal representative points $r^1, r^2$. Then $ds \leqq d(r^1, r^2) \leqq \sqrt{p}ds$. Hence, $R$ is a $ds$-uniform representation of $Y_N$.*

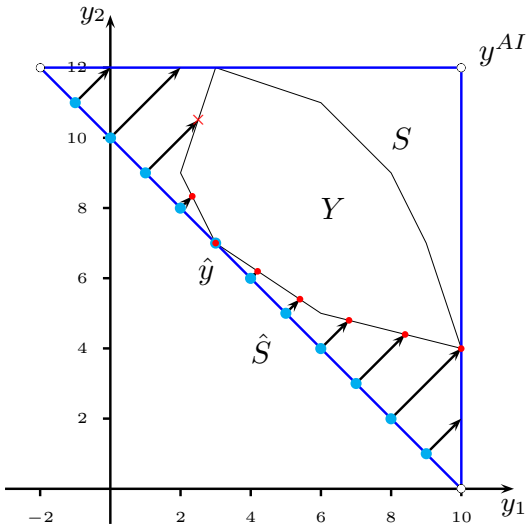



**Figure 7** The revised boundary intersection method.

2. *Assume that the width $w(S^j) \geqq ds$ for the projection $S^j$ of all maximal faces $Y^j$ of $Y_N$ on $\hat{S}$. Then $R$ is a $ds$-uniform $d_{\sqrt{p}ds}$-representation of $Y_N$.*

In this section we outline how to adapt to the situation where $Y_N$ is an explicitly given set of finitely many points. To this end, we now modify the RNBI method so that it becomes applicable to the case of $Y_N = Y = \{y^j : j \in J\}$ being an explicily given finite set. The main obstacle in doing this is that the sub-problem

$$\min\{t : q + te \in Y, t \geqq 0\}$$

that is solved for each reference pooint $q$ will most often be infeasible. To avoid this situation, we replace $Y$ in the sub-problem by $\hat{Y} = Y + \mathbb{R}^q$. Since $Y_N = \hat{Y}_N$, this has no effect on the Pareto optimal set, but the new sub-problem

$$\min\left\{t : q + te \in Y + \mathbb{R}^p_{\geqq}, t \geqq 0\right\}$$

is feasible. To solve it, we define $t(q) = \min_{j\in J} \max_{k\in\{1,\ldots,p\}}\{y^j_k - q_k\}$ and $r(q) = argmin_{j\in J}$ $\max_{k\in\{1,\ldots,p\}}\{y^j_k - q_k\}$.

To compute, for reference point $q$, the intersecion of the ray $\{q + te : t \geqq 0\}$ with the cone $y^j + \mathbb{R}^q_{\geqq}$ dominated by $y^j$, the $l_\infty$-distance $t(q)$ to the Pareto optimal point $y^j$ is computed, and the closest point to $q$ is chosen as a representative point $r(q)$. Then the representative set is $R = \{r(q) : q \in Q\}$.

There are a number of research questions related to this approach:

- Can quality guarantees in terms of uniformity and coverage error be proven?
- What is the cardinality of $R$ given the cardinality of $Q$?

### Representations based on clustering

A very simple, yet potentially effective idea for computing representations in the case of an explicitly given set of alternatives in the context of a geometric vision is based on *clustering*. The idea is to compute a certain number, say $K$ clusters very quickly and retrieve information about the quality of the representation. In many real-life datasets, the density of points is

not uniform, but has high-density clusters representing a certain "type" of outcome (e.g. products which are similar). To provide a quick overview of available Pareto optimal points, each of these types/clusters should be represented with one representative point. Clusters can be of different sizes, but would still be represented by a single point.

Such a clustering algorithm for realizing functionality 1 can be formulated as follows:

**Algorithm:** MSF-Clustering
Input: $n$ items with their objective function values; $K \in \mathbb{N}$
Output: A representation $R$ with $|R| = K$
1. Compute the pairwise distances between the items (wrt. their objective function values).
2. Sort these distances by increasing length.
3. Use Kruskal's algorithm to compute a Minimum Spanning Forest consisting of $K$ trees (=cluster).
4. For each tree: Compute median/center item as the representative point of the cluster.
5. Return all representative points.

This clustering algorithm can be implemented in a running time of $O(n^2 + n^2 \cdot \log_2 n^2 + n^2 \cdot \log_2^* n^2 + n^2) = O(n^2 \log n^2)$ and, thus, finds a representation in polynomial time. In case a DM then updates upper bounds on the values of the objectives (this is an operation which is likely to happen), a re-sorting can be implemented in $O(n^2)$ which results in an $O(n^2 \cdot \log_2^* n^2)$ algorithm for updating the representation.

Note that functionality 2, i.e. "display solutions close to some chosen representative point" can be very easily realized: all points in a cluster are displayed. Further research directions may clarify the quality of representations (wrt. uniformity and coverage) obtained with such a clustering algorithm.

### A preference-based approach

The first important question is the choice of the preference relation $\preceq$ to be used to compute the representation $R$. Relation $\preceq$ must be richer than the Pareto dominance relation in order to ensure conciseness of the representation. In cases where no a priori preference information is available, a natural candidate relation is the $\varepsilon$-*dominance* relation $\preceq_\varepsilon$ defined as follows:

$$y \preceq_\varepsilon y' \text{ iff } y_i \leq (1+\varepsilon) y'_i \qquad i = 1, \ldots, p,$$

where $\varepsilon > 0$ can be interpreted as a tolerance/indifference threshold. Note that we can use different thresholds $\varepsilon_i > 0$ for each criterion $f_i$, $i+1, \ldots, p$. We can also use additive thresholds instead of multiplicative thresholds. The relation $\preceq_\varepsilon$ enriches the standard Pareto dominance relation as illustrated in Figure 8.

In order to implement functionality 1, which aims at producing a concise representation of a region of interest, we use the concept of an $(\varepsilon, \varepsilon')$-*kernel*, introduced in [1].

▶ **Definition 3.** Given $\varepsilon, \varepsilon' > 0$, an $(\varepsilon, \varepsilon')$-*kernel* is a set of points $K_{\varepsilon,\varepsilon'} \subset Y$ satisfying:
$(i)$ for any $y' \in Y_N$ there exists $y \in K_{\varepsilon,\varepsilon'}$ such that $y \preceq_\varepsilon y'$ ($\varepsilon$-*coverage*),
$(ii)$ for any $y, y' \in K_{\varepsilon,\varepsilon'}$, not$(y \preceq_{\varepsilon'} y')$ and not$(y' \preceq_{\varepsilon'} y)$ ($\varepsilon'$-*stability*).

In order to guarantee the existence of an $(\varepsilon, \varepsilon')$-kernel, we must have $\varepsilon' \leq \varepsilon$. Considering that condition *(i)* prevails over condition *(ii)* in the definition of a good representation, we must first define a threshold $\varepsilon$ to define the precision of the representation and then set $\varepsilon'$ as large as possible. When it is possible to set $\varepsilon' = \varepsilon$, an $(\varepsilon, \varepsilon')$-kernel is called an $\varepsilon$-kernel.

Some important results, established in [1], are gathered in the following theorem.

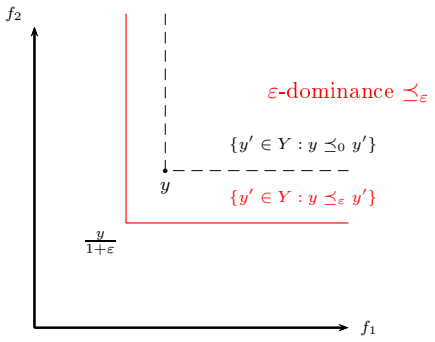



**Figure 8** Dominance ($\preceq_0$) and $\varepsilon$-dominance ($\preceq_\varepsilon$) relations.

▶ **Theorem 4.** *If $p = 2$, an $\varepsilon$-kernel always exists (with $\varepsilon' = \varepsilon$).*
*If $p \geq 3$, an $(\varepsilon, \varepsilon')$-kernel exists if and only if $\varepsilon' \leq \sqrt{1+\varepsilon} - 1$.*
*If $Y$ is defined explicitly, these concepts can be computed in a linear time.*

We show now how to implement functionality 2, which aims at producing alternatives similar to a (not necessarily) feasible reference point. Let $\bar{y}$ be the reference point. The neighborhood of $\bar{y}$ is:

$$\mathcal{N}(\bar{y}) = \{y \in Y_N : y \preceq_{\varepsilon'} \bar{y} \text{ and } \bar{y} \preceq_{\varepsilon'} y\}.$$

Note that this neighborhood is defined with a relation $\preceq_{\varepsilon'}$ which is used in the stability condition to define an $(\varepsilon, \varepsilon')$-kernel. It is indeed consistent to use this relation which was used to impose that two elements in $R$ should not be too similar. This concept is clearly computable in a linear time.

### 4.2.6 Conclusions

This report summarizes our findings on the topic of personalization of multicriteria decision support systems. With growing computational power, ever enlarging data storage capabilities, then increasing availability of large data sets and the success of multiobjective optimization methods, decision-making processes tend to ask more and more in the way of personalized aspects to make better, faster and more confident decisions. This is especially true on complex, professional applications which require sophisticated models and solution algorithms (e.g. radiotherapy treatment or landuse planning). In addition, everyday applications (such as online evaluations of products or sales for customers) with explicitly given sets of alternatives are subject to multiple criteria and a personalized perspective as well. This report identifies two central aspects which can be concisely described as "personalization in model building" for complex situations and "iterative computation of personalized representation systems" for explicitly given points. Initial ideas are presented with respect to both aspects. Yet, as the demand for personalization grows, more sophisticated concepts are still to be developed.

### References

1 C. Bazgan, F. Jamain, and D. Vanderpooten. Discrete representation of the non-dominated set for multi-objective optimization problems using kernels. *European Journal of Operational Research*, 260(3):814–827, 2017.

**2** T. Chugh, K. Sindhya, K. Miettinen, Y. Jin, T. Kratky, and P. Makkonen. Surrogate-assisted evolutionary multiobjective shape optimization of an air intake ventilation system. In *Proceedings of the 2017 IEEE Congress on Evolutionary Computation*, pages 1541–1548. IEEE, 2017.

**3** S. Dempe, G. Eichfelder, and J. Fliege. On the effects of combining objectives in multiobjective optimization. *Mathematical Methods of Operations Research*, 82(1):1–18, 2015.

**4** T. Eikelboom, R. Janssen, and T.J. Stewart. A spatial optimization algorithm for geodesign. *Landscape and Urban Planning*, 144:10–21, 2015.

**5** A. Engau and M.M. Wiecek. Cone characterizations of approximate solutions in real vector optimization. *Journal of Optimization Theory and Applications*, 134(3):499–513, 2007.

**6** A. Engau and M.M. Wiecek. Introducing nonpolyhedral cones to multiobjective programming. In V. Barichard, X. Gandibleux, and V. T'Kindt, editors, *Multiobjective Programming and Goal Programming, Proceedings*, pages 35–45. Springer, 2009.

**7** P. Eskelinen, K. Miettinen, K. Klamroth, and J. Hakanen. Pareto Navigator for interactive nonlinear multiobjective optimization. *OR Spectrum*, 23:211–227, 2010.

**8** S. Faulkenberg and M.M. Wiecek. On the quality of discrete representations in multiple objective programming. *Optimization and Engineering*, 11(3):423–440, 2010.

**9** J. Fliege. The effects of adding objectives to an optimisation problem on the solution set. *Operations Research Letters*, 35(6):782–790, 2007.

**10** M. Hartikainen and A. Lovison. PAINT-SiCon: Constructing consistent parametric representations of Pareto sets in nonconvex multiobjective optimization. *Journal of Global Optimization*, 62(2):243–261, 2015.

**11** B.J. Hunt, M.M. Wiecek, and C.S. Hughes. Relative importance of criteria in multiobjective programming: A cone-based approach. *European Journal of Operational Research*, 207(2):936–945, 2010.

**12** A.D. Jesus, L. Paquete, and J. Figueira. Finding representations for an unconstrained bi-objective combinatorial optimization problem. *Optimization Letters*, (12):321–334, 2018.

**13** S. Kaddani, D. Vanderpooten, J.M. Vanpeperstraete, and H. Aissi. Weighted sum model with partial preference information: application to multi-objective optimization. *European Journal of Operational Research*, 260(2):665–679, 2017.

**14** O. Kariv and S. L. Hakimi. An algorithmic approach to network location problems. I: The $p$-centers. *SIAM Journal on Applied Mathematics*, 37(3):513–538, 1979.

**15** K. Klamroth and J. Tind. Constrained optimization using multiple objective programming. *Journal of Global Optimization*, 37(3):325–355, 2007.

**16** K. Miettinen, M.M Makela, and T. Mannikko. Optimal control of continuous casting by nondifferentiable multiobjective optimization. *Computational Optimization and Applications*, 11:177–194, 1998.

**17** M. Monz. Pareto navigation - interactive multiobjective optimisation and its application in radiotherapy planning. *Technical University of Kaiserslautern, Department of Mathematics*, 2006.

**18** M. Monz, K.-H. Küfer, T. Bortfeld, and C. Thieke. Pareto navigation - algorithmic foundation of interactive multi-criteria imrt planning. *Physics in Medicine and Biology*, 53:985–998, 2008.

**19** V.D. Noghin. Relative importance of criteria: a quantitative approach. *Journal of Multi-Criteria Decision Analysis*, 6(6):355–363, 1997.

**20** V.D. Noghin. What is the relative importance of criteria and how to use it in MCDM. In M. Koksalan and S. Zionts, editors, *Multiple Criteria Decision Making in the New Millennium, Proceedings*, pages 59–68. Springer, 2001.

21 S.S. Ravi, D. . Rosenkrantz, and G.K. Tayi. Facility dispersion problems: Heuristics and special cases. In F. Dehne, J.-R. Sack, and N. Santoro, editors, *Algorithms and Data Structures*, pages 355–366, Berlin, Heidelberg, 1991. Springer.

22 S. Sayın. Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming. *Mathematical Programming*, 87(3):543–560, 2000.

23 B. Schulze, L. Paquete, K. Klamroth, and J. Figueira. Bi-dimensional knapsack problems with one soft constraint. *Computers & Operations Research*, 78:15–26, 2017.

24 J.I. Serna. Approximating the nondominated set of R+ convex bodies. *Technical University of Kaiserslautern, Department of Mathematics*, 2008.

25 L. Shao and M. Ehrgott. Discrete representation of non-dominated sets in multi-objective linear programming. *European Journal of Operational Research*, 255:687–698, 2016.

26 K. Sindhya, V. Ojalehto, J. Savolainen, H. Niemisto, J. Hakanen, and K. Miettinen. Coupling dynamic simulation and interactive multiobjective optimization for complex problems: An APROS-NIMBUS case study. *Expert Systems with Applications*, 41(5):2546–2558, 2014.

27 T.J. Stewart and A. Joubert. Conflicts between conservation goals and land use for exotic forest plantations in South Africa. In E. Beinat and P. Nijkamp, editors, *Multicriteria Analysis for Land-Use Management*, pages 17–31. Springer, 1998.

28 D. Vaz, L. Paquete, Fonseca C.M., K. Klamroth, and M. Stiglmayr. Representation of the non-dominated set in biobjective discrete optimization. *Computers & Operations Research*, 63:172–186, 2015.


## 4.3 Usable knowledge extraction in multi-objective optimization: An analytics and "innovization" perspective (WG3)

*Carlos A. Coello Coello, Kerstin Dächert, Kalyanmoy Deb, José Rui Figueira, Abhinav Gaur, Andrzej Jaszkiewicz, Günter Rudolph, Lothar Thiele, and Margaret M. Wiecek*



**Abstract.** Knowledge extraction aims at detecting similarities and patterns hidden in the Pareto-optimal solutions arising from the outcome of a multi-objective optimization problem. The patterns may emerge from generic relationships of several variables or objective functions. Knowledge extraction is expected to bring out valuable information about a problem and is termed as a task of "innovization" elsewhere. While certain automated innovization methods have been proposed, in this report, we attempt to formalize the overall computational task from a machine learning and data analytics point of view. The results can be used to improve modeling and understand interdependencies among different objectives.

### 4.3.1 Introduction

The topic was proposed by one of the participants (Deb) who has introduced the original idea, has been working on this topic for nearly two decades, and has called it "innovization" (innovation through optimization) [5, 6] of (technical) models, which leads to new designs, hence, true innovations.

The basic innovization idea has been used towards automated innovization methods, for example, in [2, 1, 3, 4]. The concept has been applied in practice, see, for example, [10, 13, 16]. Innovization methods have also been implemented by different other visualization or machine learning methods [14, 17, 18, 15, 7]. Since we do not aim at only reformulating the concept

of innovization but also contributing new ideas, we use the more general term "knowledge extraction". Due to the recent hype in machine learning and data analytics, this topic is of high interest. Moreover, it fits very well to the topic of this Dagstuhl seminar.

### 4.3.2 Problem statement

The main idea of knowledge extraction is as follows. Assume we have already modeled a problem with at least two conflicting objectives, that is, we have formulated a multi-objective optimization problem with certain variables, objective and constraint functions. Some of the variables might be restricted to take only discrete values which turns the problem into a mixed-integer multi-objective optimization problem. Furthermore, motivated from mechanical examples, the model contains certain parameters which are specified by the end users but might change in response to the new knowledge offered by the knowledge extraction procedure.

During the Dagstuhl seminar, we have decided to work on the following topics.

1. General Framework: Given two sets, $P$ (target set) and $Q$ (non-target set), from a problem,
   **RQ1:** What features of the problem (described by variables ($x$), objectives ($f$), inequality constraints ($g$) and equality constraints ($h$), or any other basis functions ($b$)) are present in $P$ (but not in $Q$)?
   **RQ2:** How to represent features?
   **RQ3:** How to find rules (knowledge) in a computationally efficient way?
   **RQ4:** In what ways can we utilize the "knowledge"?
2. For RQ1: We shall show some examples to clarify the description of "feature". Features to be considered will be of the type "`if condition, then decision`". The outcome for such a feature when applied to a problem vector $(x, f, g, h, b)$ is `true or false`. We shall consider problems having (i) continuous, (ii) mixed-integer, and (iii) combinatorial variables. We will refer to this as Task 1 in the following.
3. For RQ2: As Task 2, we shall identify subsets of variables defining features and show some examples. The following methods can be used:
   - User-supplied
   - ANOVA, Statistics
   - AIC, Entropy
   - Forward Selection, Backward Selection, and
   - Rough Sets.
4. For RQ3: We shall develop feature (knowledge) extraction procedures (referred to as Task 3) to find hidden features in problem instances and data using the following methods:
   - Genetic Programming (GP) to find general (free-form) features
   - Two-level Decision Tree/Forest Approach to find decision trees, and
   - Other generic or specific methods to find problem-specific structures.
5. For RQ4: We shall utilize the developed features (knowledge) to facilitate the following tasks (we refer to this as Task 4):
   - Knowledge elicitation to users in terms of (i) product platform scaling, (ii) putting focus on key concepts, (iii) using observed knowledge to build theory about knowledge, and (iv) provide leadership.
   - Online utilization of knowledge to improve convergence properties of optimization algorithms.

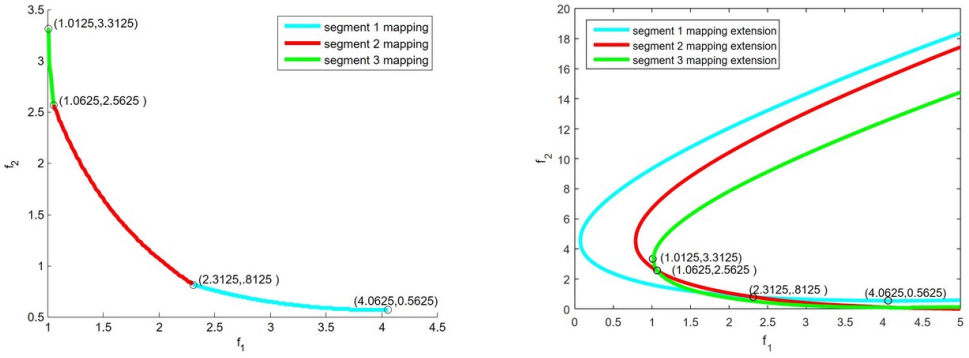



**Figure 9** The Pareto-optimal set (left) and curves determining its three nonlinear segments (right) for the BOQP (21).

- Knowledge accumulation to modify/trust the original problem in establishing the fact that (i) some variables may be redundant, (ii) some objectives may be redundant, and (iii) some constraints may be redundant.

### 4.3.3 Examples

#### Continuous optimization

As a simple illustrative example consider the biobjective quadratic problem (BOQP) in which two quadratic objective functions are minimized on a feasible set determined by three linear constraints:

$$\begin{aligned} \min_{x_1,x_2} \quad & [f_1(x_1,x_2) = (x_1-2)^2 + (x_2-2)^2, \quad f_2(x_1,x_2) = x_1^2 + (x_2-3)^2] \\ \text{s.t.} \quad & g_1(x_1,x_2) = x_1 + x_2 - 2.75 \leq 0 \\ & g_2(x_1,x_2) = 2x_1 + x_2 - 3.75 \leq 0 \\ & g_3(x_1,x_2) = x_2 \leq 2.25 \end{aligned} \tag{21}$$

The Pareto-optimal set in the objective space $(f_1, f_2)$ of this BOQP is shown in the left part of Figure 9. This set is composed of three curves whose equations, due to the simple structure of the objective functions and the feasible set, can be derived analytically. The three curves are depicted in the right part of this figure and have the following equations

$$\begin{aligned} &\text{segment 1: } f_2 = f_1 - \sqrt{16f_1 - 1} + 4.5 && \text{for } 2.3125 \leq f_1 \leq 4.0625 \\ &\text{segment 2: } f_2 = f_1 - \sqrt{18f_1 - 14.0625} && \text{for } 1.0625 \leq f_1 \leq 2.3125 \\ &\text{segment 3: } f_2 = f_1 - \sqrt{12.8f_1 - 12.96} + 2.3 && \text{for } 1.0125 \leq f_1 \leq 1.0625 \end{aligned}$$

Recalling that the Pareto-optimal set is the image of the efficient set in the decision space $(x_1, x_2)$, we observe that in this particular example each of the three curves is the image of the efficient solutions that are located along (part of) the active constraint $g_i(x_1, x_2) = 0, i = 1, 2, 3$. In effect, we obtain the rules for the decision variables $x_1$ and $x_2$ in the form $g_i(x_1, x_2) = 0$ and the answer to the research questions RQ1 and RQ2. Since in this very simple example the rules uniquely determine the Pareto-optimal set, the knowledge extraction is complete and the obtained knowledge is ultimate.

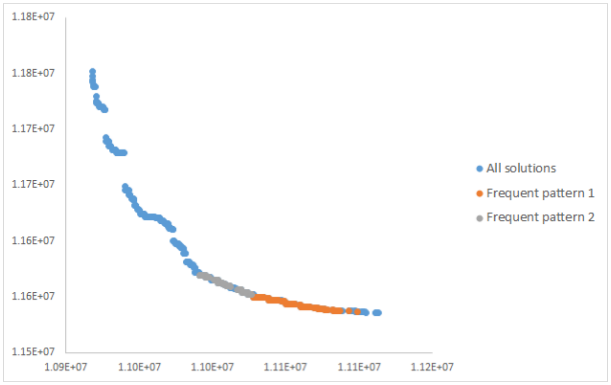



**Figure 10** Solutions in the objective space corresponding to two different frequent patterns.

## Combinatorial optimization

As another example we use a bi-objective traveling salesperson problem (TSP). We use real-life data provided by the company Emapa[1]. The data contains travel times and distances between each pair of 500 points located in Poland. The travel times and distances were estimated using data about the real-life road network. The goal is to find a Hamiltonian cycle in this graph taking into account two objectives: total travel time and total distance. The two objectives are obviously highly correlated but also partially in conflict. For example, a route utilizing highways may be longer but faster than a route using secondary roads.

To solve this problem we used a Two Phase Pareto Local Search (TPPSL) algorithm [12]. In the first phase we used the Lin-Kernighan heuristic [11] to generate an initial set of potentially Pareto-optimal solutions that are passed over to the second phase in which Pareto Local Search was run. As a result, 469 potentially Pareto-optimal solutions were found. Each solution is characterized by a set of 500 edges forming a Hamiltonian cycle. The first interesting observation is that these sets are highly similar. There are only 679 distinct edges that appear at least once in this set of solutions out of 124750 possible edges. Furthermore, 245 edges appear in all of the solutions. This set of common edges could be interpreted as a frequent pattern [9] with support (i.e., the number of matching solutions) equal to the number of all potentially Pareto-optimal solutions. In other words, each solution contains only 255 (out of 500) volatile edges, i.e., edges that do not belong to all of the Pareto-optimal solutions.

Furthermore, we can search for other interesting frequent patterns with lower support. For example, Figure 10 shows solutions supporting two different patterns presented in the objective space. The two patterns were selected such that they are supported by at least 100 solutions each, they contain many edges, and the sets of supporting solutions are disjoint. The first pattern is supported by 111 solutions and contains 346 edges. The second pattern is supported by 100 solutions and contains 367 edges. As can be seen from Figure 10, the solutions supporting the two patterns are located in different regions of the objective space. This is an interesting observation since the values of the objectives were not taken into account while selecting the two patterns. The two patterns could be understood as characterizations of two regions of the set of potentially Pareto-optimal solutions.

[1] http://emapa.pl/

### 4.3.4 Usefulness of knowledge extraction

The a posteriori analysis of the results often reveals interesting information on the problem at hand. Unfortunately, this analysis is computationally demanding, in general. Therefore, it can be critically asked what this approach serves for when the optimization process has already been terminated. In the above mentioned combinatorial example, knowledge extraction might be used to reduce the problem size by neglecting edges which never or seldom belong to Pareto optimal solutions. This might have a considerable effect on computational time while basically maintaining the quality of the solutions obtained.

Another important and useful application is in the context of online algorithms. In many applications, the same optimization problem has to be solved over and over again with only slightly different parameter values, e.g., in the context of rolling planning in energy or water networks. In these cases, knowledge extraction might help in solving subsequent optimization problems much quicker and, thus, improving solution quality tremendously when only a very short computational time for optimization is available. Moreover, if time is restricted, a true multi-objective analysis offering different Pareto-optimal solutions is typically not possible. Hence, it is of urgent interest to learn an appropriate setting of 'multi-objective' parameters quickly. Also this task can be handled well by knowledge extraction methods.

We shall work on developing methodologies for each of the above topics and plan to write a journal quality paper.

### 4.3.5 Conclusions

Pareto-optimal or near-Pareto-optimal solutions of multi-objective optimization problems often possess specific properties that can be, for example, seen from certain patterns in the variable values. Since general optimality conditions (like, for example, multi-objective variants of the KKT conditions) are often difficult to apply for practical and complex problems (for example, due to rather restrictive assumptions on the problem structure and due to the need of finding derivatives and a reliable solution of nonlinear equations and inequalities), the "innovization" procedure proposed by Deb [5] is a viable strategy. It is a two-step procedure in which first a set of preferable trade-offs and near-Pareto-optimal solutions are found by an EMO algorithm or a generative MCDM approach. In the second step, the optimized solutions are analyzed to decipher invariant features describing the variables, objectives, and constraint values that exist in the data. The usefulness of the basic innovization approach has been demonstrated by Deb and his collaborators over the past 15 years and certain efforts to automate the second step using machine learning procedures have also been proposed [2, 1, 8].

In this report, we have attempted to formalize a systematic procedure for the innovization task and to extend the basic concept to various computational, theoretical, and application domains.

## Acknowledgments

The authors acknowledge the Dagstuhl Committee for providing the excellent facility and environment that are inducive for fruitful discussions.

## References

1 S. Bandaru and K. Deb. Automated innovization for simultaneous discovery of multiple rules in bi-objective problems. In *Proceedings of Sixth International Conference on Evolutionary Multi-Criterion Optimization (EMO-2011)*, pages 1–15. Heidelberg: Springer, 2011.

2 S. Bandaru and K. Deb. Towards automating the discovery of certain innovative design principles through a clustering based optimization technique. *Engineering optimization*, 43(9):911–941, 2011.

3 S. Bandaru, A. H. C. Ng, and K. Deb. Data mining methods for knowledge discovery in multi-objective optimization: Part a – surve. *Expert Systems With Applications*, 70:139–159, 2017.

4 S. Bandaru, A. H. C. Ng, and K. Deb. Data mining methods for knowledge discovery in multi-objective optimization: Part b – new developments and applications. *Expert Systems With Applications*, 70:119–138, 2017.

5 K. Deb. Unveiling innovative design principles by means of multiple conflicting objectives. *Engineering Optimization*, 35(5):445–470, 2003.

6 K. Deb and A. Srinivasan. Innovization: Innovating design principles through optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2006)*, pages 1629–1636, New York: ACM, 2006.

7 Catarina Dudas, Amos HC Ng, Leif Pehrsson, and Henrik Boström. Integration of data mining and multi-objective optimisation for decision support in production systems development. *International Journal of Computer Integrated Manufacturing*, (ahead-of-print):1–16, 2013.

8 A. Gaur and K. Deb. Effect of size and order of variables in rules for multi- objective repair-based innovization procedure. In *Proceedings of Congress on Evolutionary Computation (CEC-2017) Conference*. Piscatway, NJ: IEEE Press, 2017.

9 Jiawei Han, Hong Cheng, Dong Xin, Xifeng Yan. Data Mining and Knowledge Discovery, August 2007, Volume 15, Issue 1, pp 55–86.

10 Nozomi Hitomi, Hyunseung Bang, and Daniel Selva. Extracting and applying knowledge with adaptive knowledge-driven optimization to architect an earth observing satellite system. In *AIAA Information Systems-AIAA Infotech@ Aerospace*, page 0794. 2017.

11 S. Lin and B.W. Kernighan. An effective heuristic algorithm for the traveling-salesman problem, Operations Research, 21, 1973, 498–516.

12 Thibaut Lust and Jacques Teghem. Two-phase Pareto local search for the biobjective traveling salesman problem, "Journal of Heuristics", 2010, Jun, 16, 3, 475–510.

13 Shigeru Obayashi, Shinkyu Jeong, and Kazuhisa Chiba. Multi-objective design exploration for aerodynamic configurations. *AIAA Paper*, 4666:2005, 2005.

14 Shigeru Obayashi and Daisuke Sasaki. Visualization and data mining of pareto solutions using self-organizing map. In *Evolutionary multi-criterion optimization*, pages 796–809. Springer, 2003.

15 Akira Oyama, Taku Nonomura, and Kozo Fujii. Data mining of pareto-optimal transonic airfoil shapes using proper orthogonal decomposition. *Journal of Aircraft*, 47(5):1756–1762, 2010.

16 Kazuyuki Sugimura, Shigeru Obayashi, and Shinkyu Jeong. Multi-objective design exploration of a centrifugal impeller accompanied with a vaned diffuser. In *ASME/JSME 2007 5th Joint Fluids Engineering Conference*, pages 939–946. American Society of Mechanical Engineers, 2007.

17 H Taboada and D Coit. Data mining techniques to facilitate the analysis of the pareto-optimal set for multiple objective problems. In *Proceedings of the 2006 Industrial Engineering Research Conference (CD-ROM)*, 2006.

18 Tamara Ulrich, Dimo Brockhoff, and Eckart Zitzler. Pattern identification in pareto-set approximations. In *Proceedings of the 10th annual conference on Genetic and evolutionary computation*, pages 737–744. ACM, 2008.

## 4.4 Complex networks and MCDM (WG4)

*Richard Allmendinger, Michael Emmerich, Georges Fadel, Jussi Hakanen, Johannes Jahn, Boris Naujoks, Robin Purshouse, and Pradyumn Shukla*



### 4.4.1 Introduction

This report introduces a novel, generic, multi-layered network model of large-scale multi-criteria decision making (MCDM), with a focus on the design and optimization of complex products and platforms. The report provides some examples of network structures in MCDM applications and develops two use-cases for the multi-layered model.

The starting point for this model was to view MCDM from the perspective of complex networks. The study of complex networks (CS) is a topic that recently received considerable attention across various disciplines [2]. Typical questions investigated in CS are:

- modeling/formalization and visualization of networks;
- dynamics of networks – both in terms of states and in terms of structure;
- microscopic (e.g., node degrees, node centrality) and macroscopic (e.g., moments of the degree distribution, sparsity, modularity, community structure) properties of networks and how they influence each other;
- algorithms on complex networks.

### 4.4.2 Related work

In several publications, the modeling of design and optimization processes in terms of networks has been addressed. Here we provide only a snapshot of the current state of the art in this domain.

- Martins and Lambe [7] view networks by means of a matrix approach. Their focus is on the coupling between disciplines via shared design variables. The coupling matrix can be exploited by gradient based techniques via the chain rule and leads to efficient methods with sparse matrices. From a networks perspective, the matrices can be interpreted as adjacency matrices and therefore a translation into a network model is possible. However, it can be argued that the approach has a too strong focus on computational models to capture the entirety of a production environment, with aspects such as platforming, discipline specific decision making and multi-objectivity within subdisciplines. This is why we aim for a network model with a broader scope and emphasizing on linkage aspects, albeit less focused on quantitative aspects.
- Maulana et al. [6] introduce network models to model the relationships between objective functions in many objective optimization. Positive links indicate complementary objectives, negatively weighted links indicate conflicting objectives, whereas the non-existence of links signals that objectives can be optimized independently (for instance, because they depend on disjoint variable sets). They propose a method to derive these conflict graphs empirically from a correlation matrix and use the networks to decompose the problem, detect communities of objectives, and the relationship between these communities. Despite its usefulness in structuring many-objective optimization problems, the model by Maulana et al. is limited in scope. It represents only a single layer of the multidisciplinary problem – the objectives layer - and falls short in terms of modeling and integrating relationships between design variables, subsystems, and disciplines (or decision makers).

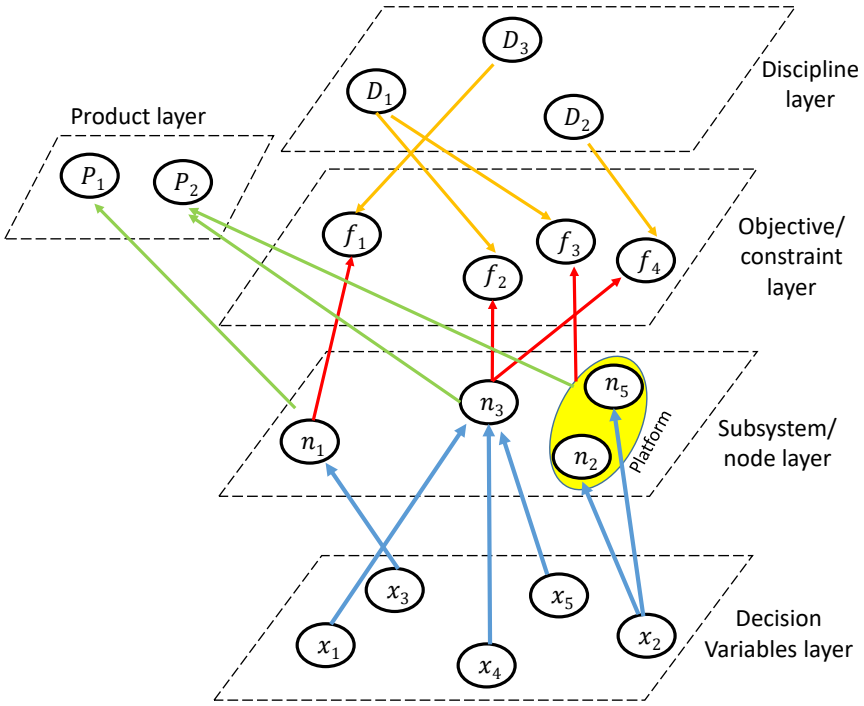



**Figure 11** Layered graph of a multidisciplinary optimization problem.

- Braha et al. [1] aims for models of a design department of a large enterprise representing interactions between designers and engineers as links. The model is not very detailed in terms of node and link semantics, but due to the large size of the data sets some interesting conclusions can be drawn about the general structure of the network, such as, scale-free degree distributions and small word properties.
- Ríos-Zapata et al. [9] consider complex decision networks in the context of traceability within product design processes. The work is preliminary in nature, but is interesting in its use of *traceability trees* to represent the multi-level relationships that connect high-level requirements to detailed design realisations. It is possible to reformulate this approach in MCDM terminology, where *Properties* are criteria, objectives or constraints ($\mathbf{f}$), *Characteristics* are decision variables ($\mathbf{x}$), *External Conditions* are parameters ($\mathbf{p}$) and *Relations* are the models (simulations or expert opinions) that map $\mathbf{x}$ to $\mathbf{f}$. The authors demonstrate the approach on a simple design problem (a portable cooler), highlighting the impact of detailed design choices on the properties of the product.
- Klamroth et al. [5] introduce the concept of *interwoven systems* for multi-objective optimization, in which design, optimization and decision making activities take place within the context of interacting sub-systems. Each sub-system can be viewed as a node within the global problem, with edges that represent shared variables and dependencies. The paper is particularly notable in developing Pareto optimality definitions for such interwoven systems.

### 4.4.3 Towards a formalisation of complex networks for MCDM

Taking inspiration from the approach of Ríos-Zapata et al. [9], a multiobjective optimization problem in a multidisciplinary product design setting can be viewed as a layered graph, consisting of layers. The layered graph is visualized in Figure 11. Each layer has its own specific type of nodes.

**$L_1$** Elementary Variables: The nodes of this layer are the decision variables. One might also include the environmental variables which cannot be controlled.

**$L_2$** Subsystems: This layer consists of the subsystems of the production process. Subsystems have often their individual modeling and simulation approaches that are then combined in the multidisciplinary optimization.

**$L_3$** Objectives and Constraints: This layer consists of objectives and constraints. Some constraints and objectives are formulated across different subsystems. An example is the total mass of a car, to which different subsystem designs contribute, such as engine and chassis, but others not, such as navigation software.

**$L_4$** Disciplines: Disciplines are concerned with different aspects of the design. For instance, in car design, one might think of aerodynamics, car electronics, product marketing, and engine design. Typically, in a product design process, disciplines are represented by different teams with their own specific responsibilities. They are concerned with specific objectives and constraints. For instance, the aerodynamics of a car might be of concern for the marketing and for the environmental efficiency of a car.

**$L_5$** Products: the products are introduced into the model, in order to model platforming strategies. A platform is a hyperedge of subsystems that can be produced in a combined way and enter in this way into product. As a platform might include more than two nodes, hyperedges (subsets of nodes) are considered as a model.

Relationships occur between nodes of different layers. Aiming for not modeling relationships that can be deduced by means of transitive closure, we model only relationships of the following types:

**$E_{12}$** Variable, Subsystem relationships: Subsystems can be viewed as functions that map decision variables to outputs, that are then used to compute objective and constraint function values. Formally, $E_{12} \subseteq L_1 \times L_2$.

**$E_{23}$** Subsystems, Objectives relationships: The behavior and properties of subsystems contribute to some of the objectives and constraints. Formally, $E_{23} \subseteq L_2 \times L_3$.

**$E_{34}$** Objectives, Disciplines relationships: Disciplines take into consideration certain objectives and constraints, and it is possible that objectives and constraints are shared among multiple disciplines. Formally, $E_{34} \subseteq L_3 \times L_4$.

**$E_{25}$** Subsystem/Platform, Products relationships: Products consist of subsets of subsystems, that might be grouped to subsets (platforms). Formally, $E_{25} \subseteq \wp(L_2) \times L_5$. Here $\wp(L_2)$ denotes the set of potential platforms (subsets of subsystems with cardinality bigger than 1) and subsystems, represented by the singletons. Non-overlap in terms of subsystems applies.

All relationships are many-to-many relationships. There is total participation of each node set in the relationship sets. This is visualized in Figure 12. An overview of the components of a complex network for MCDM can also be found in Table 3.

### 4.4.4 Examples of complex networks

The above framework can be used to represent a range of applications comprising interconnected components, which may be product parts and parameters, decision makers, and/or objectives. Examples of applications that would fit the framework include the design of a

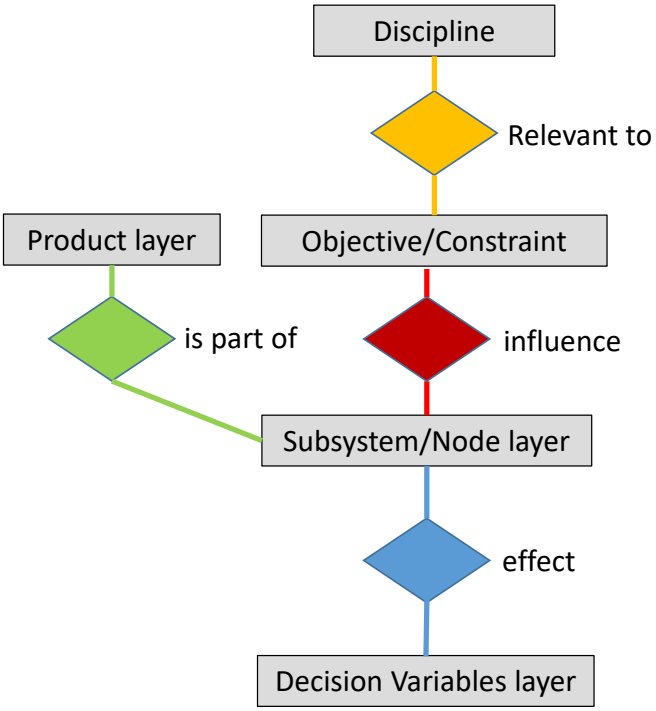



**Figure 12** Entity Relationshup Diagram.

product, planning and constructing of a facility, and determining the optimal location of facilities (e.g. location of a school). The next two sections look in more detail on how the framework can be mapped to two particular applications: designing a car (Section 4.4.4) and planning and constructing a power station (Section 4.4.4).

### Example I: Product design

Being able to model and facilitate the complex process of designing a sophisticated product was one of the main motivations of this work. The reason that this task is not straightforward is that a complex product, such as a car, consists of a large number of interconnected components, such as an engine, the car body, suspension, electrical supply system, etc. as illustrated in Figure 13. These components are developed by different teams often independently of each other and with more or less conflicting goals in mind.

The availability of a structured framework to support the design of a complex product, such as a car, will make the design process more efficient and cheaper as well as provide a tool to visualize to the entire design team the various design components and their relationships. Ultimately, the framework will facilitate decision making in an environment that exists of many decision makers and different (conflicting) design goals (objectives).

What follows is a layer by layer mapping of the framework to the process of designing a car (a less formal mapping is carried out in the next example).

**Decision variables layer:** This layer comprises controllable parameters that have a direct influence on the shape, size and operation of every single component of a product. In the car design example, this may include appearance parameters, such as the dimensions of

| Nodes | Edges | Layer |
|---|---|---|
| Disciplines or Discipline Decision Makers | Objectives and Constraints relevant to the discipline/decision making | Discipline |
| Objectives and Constraints | Nodes and Subsystems that influence objectives | Objective/ Constraint |
| Subsystems and Nodes, can be grouped to platforms | Variables that effect the subsystems | Subsystem/ Node |
| Elementary variables | controllable/observable | Decision Variable |

**Table 3** Overview of components, i.e. nodes, edges, and layers, of a complex network for MCDM.

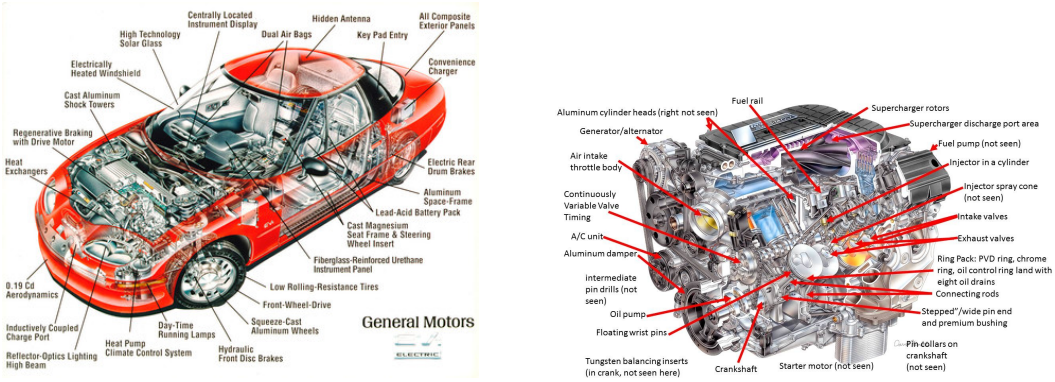



**Figure 13** Car parts (left plot, source: www.pinterest.co.uk) and engine parts (right plot, source: www.anatomybody101.org).

a component and the location of a component within the overall product, and operation parameters, such as mass, energy, power, temperature, etc required to run a particular component.

**Subsystem / node layer:** The decision variable layer feeds into the subsystem layer in the sense that specifying the setting of each of the decision variables will define the appearance and working of a subsystem or component, such as the engine, battery, suspension, chassis, and car body. Each component has objectives and constraints (e.g. related to the power and noise of engine, and weight of chassis), which need to be accounted for when setting the decision variables. Typically, there is a decision maker for each subsystem (component) aiming to get the component at hand as optimal as possible.

The combination of several components can make up a platform. For instance, in the context of cars, the combination of engine and suspension type may define a platform. The characteristics of a platform, such as size and components involved, are monitored and decided by a decision maker, who is typically different from the ones governing the components involved in a platform.

**Disciplines layer:** This layer sits above the subsystem layer because it addresses multiple components to best satisfy a joint objective and meet certain constraints. Examples of disciplines in car design include the acoustics of a car (noise), structures, dynamics, aerodynamics, and heat transfer. It is common that each discipline has a decision maker associated with it.

**Figure 14** Combined cycle power plant at Düsseldorf, Germany (source: www.siemens.com/press).

**Product layer:** The product layer links multiple components and platforms to form the overall physical product, e.g. an actual car.

**Objective / constraint layer** All objectives and constraints considered in the subsystem layer and the discipline layer are mapped onto the objective / constraint layer. In general, there are several (conflicting) objectives in that layer including obvious ones, such as costs, but also several other objectives one needs to account for prior to rolling out a product, such as manufacturability, environmental impact, sustainability, product robustness, and customer satisfaction. These objectives are typically posed by the chief engineer.

### Example II: Power station planning and construction

Planning and construction of power stations is a very difficult task, which leads to a very complex network with various complicated subsystems. Each power station is a personalized product (*product layer*) because it is designed to the specific requests of every customer. Various technologies (*subsystem layer*) are possible – the so-called combined heat and power plant is the most cost-efficient way to produce power and heat (e.g., compare Figure 14).

In this subsection we present and discuss results obtained by Hirschmann [3, 4]. There are several stages of the resulting engineering process including first planning, tender compiling, assembling and integration, putting into operation and service. Since this large problem is a discrete-continuous multiobjective optimization problem, the *variables layer* consists of real variables (e.g. duration times of subprocesses), integer variables (e.g. number of stuff members) and attributes (e.g. describing the quality of tools). The *objective layer* considers five objectives: Project costs, fixed costs and duration have to be minimized, and flexibility and the effective use of the resources are to be maximized. Figure 15 illustrates the *discipline layer* together with the cooperation of the fields. Besides the classical disciplines such as electrical engineering and mechanical engineering, there are additional disciplines, such as fire and noise protection, among others.

The resulting optimization problem leads to an optimal engineering process illustrated in Figure 16. This optimal process also considers time frames and is given in a simplified form. Several tasks are done in parallel, but there are also common nodes as a result of these subprocesses.

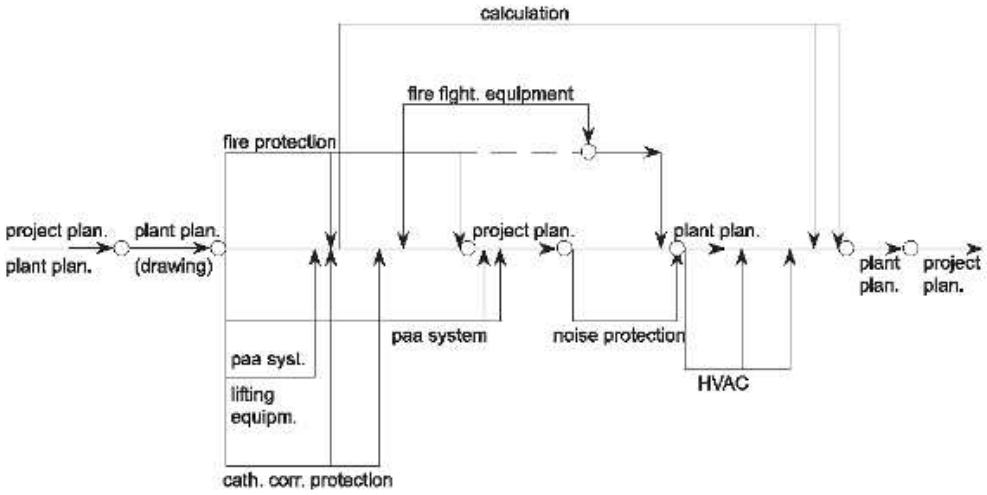



**Figure 15** Discipline layer of the construction of a power station.

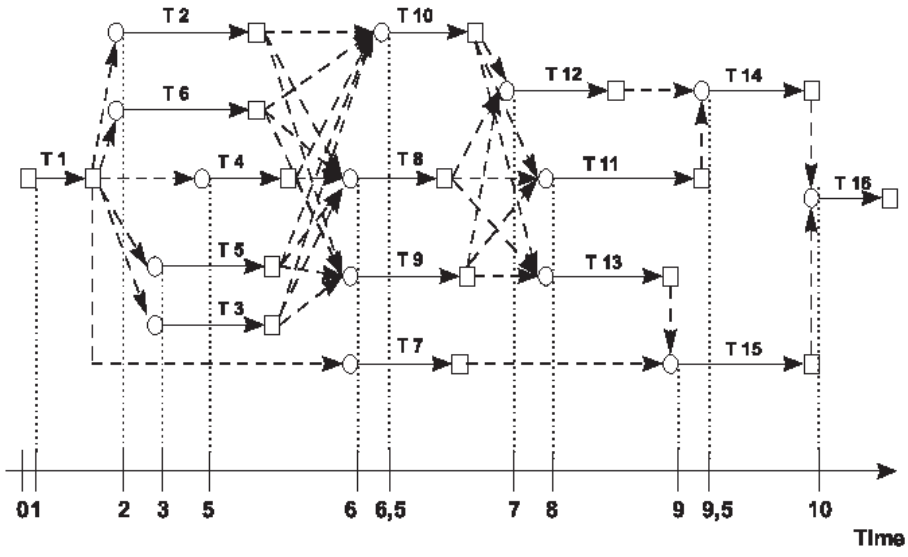



**Figure 16** Optimal engineering process of the construction of a power station.

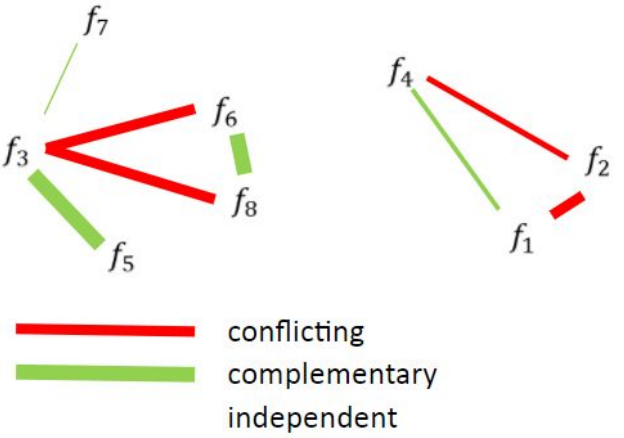



**Figure 17** An example of empirical correlations between objectives.

### 4.4.5 Use cases for decision making in complex networks

Based on the above framework of complex networks, decision making is happening in different layers having decision makers with roles as described in examples of Section 4.4.4. Most of these decision situations are multi-objective by nature and the objectives in a lower level are typically a subset or a part of the objectives considered in higher levels which makes decision making in this setting complex. Next we will present some possible use cases for supporting multiple criteria decision making in complex networks.

#### Identifying conflict and redundancy

One use case is to use empirical correlations at objective layer to identify relationships between objectives, i.e. conflicting, harmonious and independent objectives [8]. When proceeding downwards, one can identify candidates for platforms that minimize potential conflict(s). In other words, what is the set of subsystems that can be used as a platform common to different products that minimizes potential conflicts. To evaluate this, new metric(s) are needed. On the other hand, when moving upwards, decision hotspots can be identified. That means identifying decision makers / disciplines with conflicts of interest related to the objectives considered requiring communication and negotiation in order to find consensus. Finally, within the objective layer, empirical correlations can be used to find and remove redundant objectives. An example of empirical correlations is shown in Figure 17 where green color denotes positive correlation while red color indicates conflicts.

#### Case-based reasoning for product design programmes

A further potential use case for a complex MCDM network is the ability to identify likely sources and degrees of conflict within product and platform design programmes. If existing product design programme exemplars can be captured using layered graphs, then network statistics can be used to quantify the features of these processes. For existing and past programmes, experiential design expertise is often available on the presence of conflict within the programme. This combined evidence could be used to develop case-based reasoning for new product design programmes, indicating the likely levels of conflict that will be experienced in the design of the product. This intelligence could be used by organisations in resource planning and management for forthcoming design programmes.

### 4.4.6 Discussion and future research ideas

We originally started our work on complex networks and MCDM collecting research questions that came up thinking about the topic. Here are some of the research questions we discussed:

1. What examples of complex MCDA networks exist?
2. Can we simplify these to tractable examples?
   - What is minimum representation of multi-objective decision problem?
3. How do we represent these using formal languages?
4. How do we incorporate platform design issues?
5. How can we characterise the networks?
6. How do we analyse, design, optimise (on) these networks?
7. How do we introduce platforms in the networks?
8. How do we support decision making/consensus building on the networks?
9. What questions do we want to ask the network:
   - Who/what are the critical components wrt consensus finding?
   - Can we define useful metrics?
     These might be uniqueness, computability, resilience, conflicts (levels and causes) etc.

During our work on the topic, we have been able to find some answers to these questions. For example, Section 4.4.4 provides two examples of complex MCDM networks as an answer to question 1. In addition, our attempt to define use cases can be seen as an answer to question 2, however, not considering the minimisation aspect raised in the subquestion. Finally, question 3 was the starting point for Section 4.4.3 on formalisation. Answering the remaining question is part of future work.

#### References


1 Dan Braha and Yaneer Bar-Yam. The statistical mechanics of complex product development: Empirical and analytical results. *Management Science*, 53(7):1127–1145, 2007.

2 Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5):175–308, 2006.

3 Claudia Hirschmann. Optimierung von Engineering-Prozessen. *PhD thesis, University of Erlangen-Nürnberg*, 1998.

4 Claudia Hirschmann. Optimization of engineering-processes. *Poster*, 1998.

5 K. Klamroth, S. Mostaghim, B. Naujoks, S. Poles, R. Purshouse, G. Rudolph, S. Ruzika, S. Sayın, M.M. Wiecek, and X. Yao. Multiobjective optimization for interwoven systems. *Journal of Multi-Criteria Decision Analysis*, 24(1–2):71–81, 2017.

6 Asep Maulana, Zhongzhou Jiang, Jing Liu, Thomas Bäck, and Michael TM Emmerich. Reducing complexity in many objective optimization using community detection. In *Evolutionary Computation (CEC), 2015 IEEE Congress on*, pages 3140–3147. IEEE, 2015.

7 Joaquim RRA Martins and Andrew B Lambe. Multidisciplinary design optimization: a survey of architectures. *AIAA journal*, 51(9):2049–2075, 2013.

8 Robin C. Purshouse and Peter J. Fleming. Conflict, harmony, and independence: Relationships in evolutionary multi-criterion optimisation. *Lecture Notes in Computer Science*, 2632:16–30, 2003.

9 David Ríos-Zapata, Jérôme Pailhes, and Ricardo Mejía-Gutiérrez. Multi-layer graph theory utilisation for improving traceavility and knowledge management in early design stages. In *Procedia CIRP 60*, pages 308–313, 2017.

## 4.5 Meta-modeling for (interactive) multi-objective optimization (WG5)

*Dimo Brockhoff, Roberto Calandra, Manuel López-Ibáñez, Frank Neumann, Selvakumar Ulaganathan*



### 4.5.1 Introduction

An important factor in evaluating multi-objective optimization (MOO) algorithms is data efficiency. For many real-world optimization problems, the number of evaluations of the objective functions that can be performed is limited due to cost, time or system constraints. Therefore, it is paramount for future MOO algorithms to be as data efficient as possible.

One approach for improving data efficiency is the use of meta-modeling (also called in different communities Kriging or Bayesian optimization). The underlying idea behind meta-modeling approaches is that explicitly building a model from the data collected during the optimization makes possible to use this data to efficiently reason about the next set of variables to evaluate. Moreover, the use of appropriate probabilistic meta-models makes the optimization more resilient to the stochasticity of the objectives.

In the context of meta-modeling in MOO, there are several open questions that apply also to an interactive context. One fundamental question, and the one that we discuss in this report is: What meta-model should be learned? Or akin: At which level of abstraction should we create the meta-models? In the literature, we can find meta-models at different levels of abstractions: (*1*) meta-models of the the multiple objective functions by means of a model per function [4, 9], (*2*) meta-model of the value of a scalarizing function [7, 2] that is defined in terms of some weights that need to be varied at runtime to approximate the whole Pareto front or (*3*) meta-models that predict some quality metric used by the optimization algorithm, for example, the Pareto ranking of solutions [8]. For detailed citations and discussion of related work, we refer to the recent review by Horn et al. [6]. However, despite these works, it is unclear which choice is preferable under different circumstances. Moreover, among the optimization algorithms that employ the third option, we did not find any work that directly model Pareto compliant quality metrics such as the hypervolume or epsilon metrics [14].

Many recent multiobjective optimizers employ the hypervolume indicator to measure the quality of the current solution set due to its advantageous theoretical properties [14, 11]—among them also some prominent model-based algorithms [4, 9]. Although previous works in the literature have shown that it is possible to compute the expected improvement of the hypervolume contribution directly from the meta-models of the objective functions [5], we did not find any work that has attempted to model directly the hypervolume contribution of a point in the decision space.

In this work, we discuss several advantages of modeling the hypervolume contribution, provide several alternative approaches for doing so, and present preliminary numerical results. Moreover, our idea of directly modeling the hypervolume contribution can be extended to other Pareto-compliant quality metrics [14], such as the quality metric guiding IBEA [12], which is based on the binary $\epsilon$-metric.

Another important question is which is the best meta-modeling technique to use for each case. However, we hypothesize that, in the context of (stochastic) MOO, the actual technique used has probably less impact than what is actually modeled. Thus, in this report, we focus on Gaussian processes (GP) [10] and we leave the use of other meta-models for future work.

### 4.5.2 Bayesian multi-objective optimization (BMO)

Let us assume the following algorithmic framework applied to the classical Bayesian optimization scenario. We have a decision space that is a subset of $\mathbb{R}^n$, where $n$ is the number of decision variables and a vector of $M$ (expensive) objective functions $\vec{f}(\vec{x}) = \{f_1(\vec{x}), \ldots, f_M(\vec{x})\}$, where $f_i \colon \mathbb{R}^n \to \mathbb{R}$ $(1 \leq i \leq M)$, that are, without loss of generality, to be minimized. The optimization goal is to approximate, as well as possible, the Pareto-optimal set, that is the set of solutions $X^*$ that are not dominated by any other feasible solution, that is, $\vec{x}^* \in X^*$ iff $\nexists \vec{x} \in \mathbb{R}^n$ such that $\vec{f}(\vec{x}) \preceq \vec{f}(\vec{x}^*) \wedge \vec{f}(\vec{x}) \neq \vec{f}(\vec{x}^*)$, where $\preceq$ is the weak Pareto dominance relation.

A possible Bayesian optimization algorithm for solving the above problem is shown in Algorithm 1. In this algorithm, we assume that there is a method for generating an initial set of solutions available. The algorithm evaluates a single point $\vec{x}_t$ per iteration $t$ on the true vector of objective functions $\vec{f}(\vec{x}_t)$. Then, it builds a surrogate model $\mathcal{M}$ based on the set of solutions evaluated up to the last iteration $t$, $X_t = \{\vec{x}_1, \ldots, \vec{x}_t\}$ and their true objective function values $Z_t = \{\vec{f}(\vec{x}_1), \ldots, \vec{f}(\vec{x}_t)\}$. How to build the model or what is the function (or functions) predicted by the model are left unspecified. As mentioned in the introduction, these may be the individual objective functions $(f_i)$, the value of some weighted aggregation (scalarization) of the objective functions, or some quality metric applied to the image of the solution set $X_t$. The model is then exploited at each iteration to suggest the next single solution $\vec{x}_{t+1}$ to be evaluated on the true objective functions $\vec{f}$. Again, how the model is exploited depends on the particular implementation of this algorithmic model.

**Algorithm 1** Template for Bayesian Multiobjective Optimization (BMO)

1: Initially, a set of $\mu$ solutions $X_\mu = \{x_1, \ldots, x_\mu\} \in \mathbb{R}^n$ is generated by means of random sampling, Latin Hypercube Design or some other method
2: Compute $Z_\mu$, the image of $X_\mu$ by evaluating the vector of true objective functions $\vec{f}(x_i) \in \mathbb{R}^M$ for each $x_i \in X_\mu$
3: Set the iteration counter $t$ to $\mu$ (the number of so-far evaluated solutions)
4: **repeat**
5: Build a model $\mathcal{M}$ based on $X_t$ and $Z_t$
6: Use $\mathcal{M}$ to suggest a new point $\vec{x}_{t+1}$ based on an acquisition function (e.g., expected improvement)
7: Evaluate the true $\vec{f}(\vec{x}_{t+1})$ and set $X_{t+1} = X_t \cup x_{t+1}$ and $Z_{t+1} = Z_t \cup \vec{f}(\vec{x}_{t+1})$
8: **until** happy or running out of time

### 4.5.3 A surrogate model for the HV contribution

The goal of several highly effective multi-objective optimization algorithms is to maximize the hypervolume of the set of solutions found. The hypervolume of a solution set $X = \{\vec{x}_1, \ldots, \vec{x}_t\}$ $(\vec{x}_i \in \mathbb{R}^n, \forall i = 1, \ldots, t)$, given a reference point $\vec{r} \in \mathbb{R}^M$ is the hypervolume of the objective space dominated by the solution set $X$ and bounded above by the reference point:

$$HV(X) = \int \mathbf{1}_{\{\vec{z} \in \mathbb{R}^M \mid \exists \vec{x} \in X : \vec{f}(\vec{x}) \preceq \vec{z} \preceq \vec{r}\}}(\vec{z}) d\vec{z} \tag{22}$$

where $\preceq$ is the weak Pareto dominance relation and $\mathbf{1}_A(a)$ the indicator function, giving one if and only if $a \in A$. The Pareto-optimal set has the largest hypervolume of all feasible sets.

A way to guide the selection (or removal) of solutions during optimization is to select (or discard) solutions with the highest (resp. lowest) hypervolume contribution to the current

solution set, where the hypervolume contribution ($HVC$) of a solution $\vec{x}$ to a solution set $X$ is the increment in hypervolume after the addition of $\vec{x}$ to $X$, that is, $HVC(\vec{x}, X) = HV(X \cup \vec{x}) - HV(X)$. If $\vec{x}$ is dominated by any solution in $X$, then $HVC(\vec{x}, X) = 0$.

In the context of Bayesian optimization, a previous work [5] has shown that it is possible to model the true objective functions and compute the hypervolume contribution of a solution directly from this model. Our proposal here is to model directly the hypervolume contribution (or some function related to this contribution) of a point in decision space, relative to the archive $X_t$ of already evaluated solutions (and their image $Z_t$), without building any model of the actual objective functions. One motivation for modeling directly the hypervolume contribution is that we would model a single "function" instead of $M$ objective functions. Another motivation is that we conjecture that the landscape of the hypervolume contribution is likely to be more regular and easier to navigate and model than the combined landscape of the true objective functions.

To motivate this conjecture, we show in Fig. 18, for the simple problem of optimizing two Sphere functions with two decision variables, the hypervolume contribution of each point of the decision space with respect to a solution set of five solutions (marked with $\times$). The center (hence, optimal) solution of each Sphere function is marked with a red and blue point, respectively. As shown by the plot, the hypervolume contribution is a multi-modal function but looks globally well-behaved with locally quadratic shapes and in the specific case of the five given solutions, a single global optimum with a large basin of attraction. Although this is not enough to prove our conjecture, specially when we move to higher dimensions and more complex problems, it does show that the landscape of the hypervolume contribution is not necessarily more complex than the combined landscape of the actual objective functions being optimized.

In order to build a model that predicts the hypervolume contribution of each solution in the decision space, we need to find out a way to build such a model using the information contained in our current solution set $X_t$ and its image $Z_t$. We cannot simply use the hypervolume contribution of each point in $X_t$ with respect to itself, since all solutions would have zero value. We discuss several possibilities in the next subsections.

### Method 1: Use information only from dominated solutions

A first approach is to keep the assumption that the hypervolume contribution of each point nondominated with respect to the current set $X_t$ is zero, but assign a negative value to those points from $X_t$ that are dominated. We have devised up to three different ways of doing the latter, which are illustrated in Fig. 19:

**(a)** A first variant assigns $HV(\{\vec{x}\}) - HV(ND_t)$ to each dominated point $\vec{x} \in X_t$ where $ND_t$ is the set of non-dominated points in $X_t$. The main advantage of this method is its simplicity. However, this variant is not smooth around zero when $\vec{x}$ gets closer to the non-dominated set.

**(b)** A second variant assigns $HVC_{(-)}(\vec{x}, \mathrm{ND}_t))$ to each dominated point $\vec{x}$, where $HVC_{(-)}$ denotes the contribution of $\vec{x}$ over the non-dominated set $\mathrm{ND}_t$ if we maximize instead of minimizing the objective functions and using the ideal of $\mathrm{ND}_t$ as the reference point for computing the hypervolume. We call this function, *negative hypervolume contribution*.

**(c)** Another possibility is to use a distance metric from $\vec{x}$ to $\mathrm{ND}_t$.

**(d)** Another simple strategy (not shown in Fig. 19) is to assign the negative dominance rank (from non-dominated sorting) to each dominated point.

We expect that the metrics above would be able to approximate the hypervolume contribution of solutions dominating the current solution set by exploiting the inherent

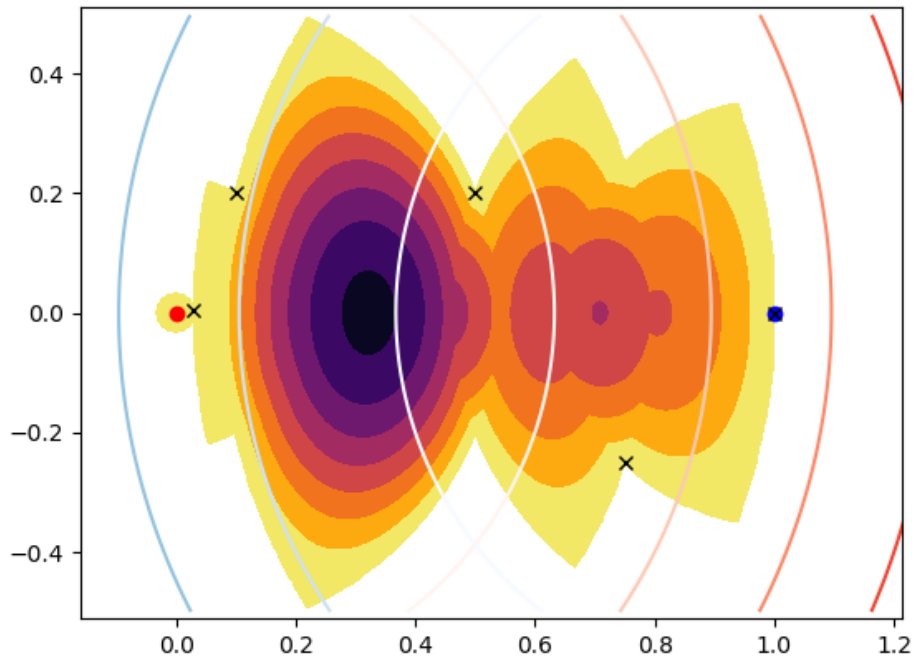



**Figure 18** Value of the hypervolume contribution of each point of the decision space of $\mathbb{R}^2$ with respect to the solution set denoted by the symbols $\times$, when minimizing two Sphere functions whose optimal solutions in $(0.0)$ and $(1, 0)$ respectively correspond to the blue and red points. Darker colors indicate larger values of the hypervolume contribution.

symmetries of meta-models such as Gaussian processes. However, the fact that no distinction is made for the nondominated solutions in our solution set may hinder the prediction power of such a model (and waste useful information). Thus, we propose next a way to assign a value to such nondominated solutions.

### Method 2: Use information from all solutions evaluated

The idea underlying our second proposed approach is to distinguish between points in the non-dominated set $ND_t$ by assigning different values to each of them (instead of zero like in our first method above) in order to give even more information to the model. In particular, given a nondominated solution $\vec{x} \in ND_t$, we assign it its actual hypervolume contribution to the set $X_t$ as $HV(X_t) - HV(X_t \setminus \vec{x})$. This should result in a model with higher values around solutions that are isolated in the objective space with the goal to force the Bayesian optimizer to suggest new solutions that are more likely to dominate a larger part of the objective space.

In the case of dominated points, we can use any of the variants discussed for method 1 above (see Fig. 19), leading to variants 2a, 2b, 2c, and 2d. Additionally, we could simply assign a value of zero for such points and only use the information provided by the nondominated solutions.

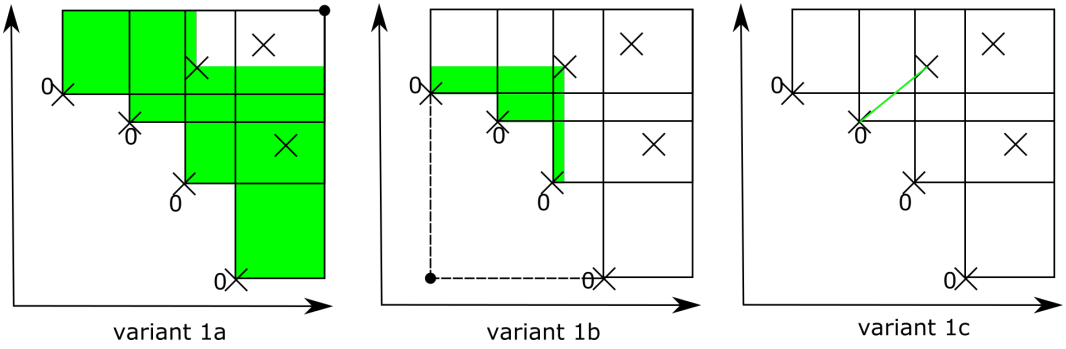



**Figure 19** Three ways of assigning an $HVC$-related value to the leftmost of three dominated solutions.

### Preliminary experiments

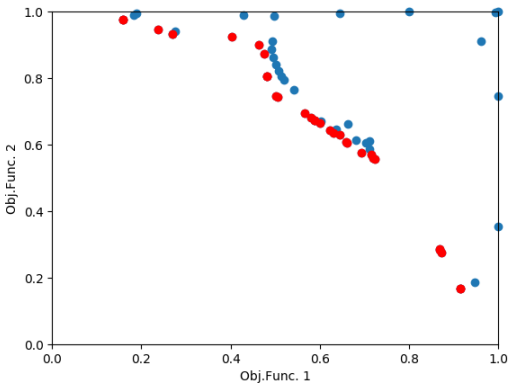
We carried out a few preliminary experiments to see whether the proposals above are able to guide optimization. In particular we analyze methods 1a and 2a. In method 1a, each dominated point $\vec{x} \in X_t \setminus ND_t$ is assigned the value $HV(\{\vec{x}\}) - HV(ND_t)$ $\vec{x} \in X_t$, where $ND_t$ is the set of non-dominated points in $X_t$, while points in $ND_t$ have a value of zero. In method 2a, dominated points have the same value as in method 1a, but each nondominated solution $\vec{x} \in ND_t$ is assigned its actual hypervolume contribution to the set $X_t$ as $HV(X_t) - HV(X_t \setminus \vec{x})$.

We prototyped and integrated these two methods in the Algorithm 1 using the Opto framework [1]. Subsequently, we executed the algorithms for a maximum of 60 evaluations of the true objective function vector. We compare the results with the well-known ParEGO [7]. Figure 20 show the objective vectors of the final solution set produced by each approach on the bi-objective Double Sphere problem with dimension $n = 2$. Nondominated solutions are shown in red, while dominated solutions are shown in blue. The caption below each plot indicates the size and the hypervolume of the nondominated set produced by each approach. Although ParEGO produces the best results, it is encouraging that the first two runs of our proposed approaches produce reasonable results. In particular, Method 2a produces slightly better hypervolume but seems to have trouble generating solutions in the extremes of the Pareto frontier and it generates solutions that are too close to each other.

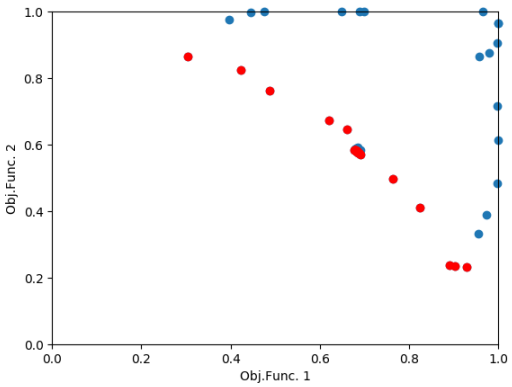
When looking at the solution space (Fig. 21), we can clearly see that the solutions produced by ParEGO are well-distributed along the Pareto set (green line), whereas the solutions produced by methods 1a and 2a are clustered in a smaller region. This suggests that the meta-model predicting the hypervolume contribution is not able to find extreme solutions and keeps predicting a high hypervolume contribution in that small region.

We also apply the three approaches to the more challenging ZDT1 problem [13] and results are shown in Fig. 22. To our surprise, our two methods are able to obtain slightly higher hypervolume values than ParEGO, although only method 2a shows an even distribution of solutions along the Pareto frontier, whereas method 1a produces solutions clustered in two small regions.

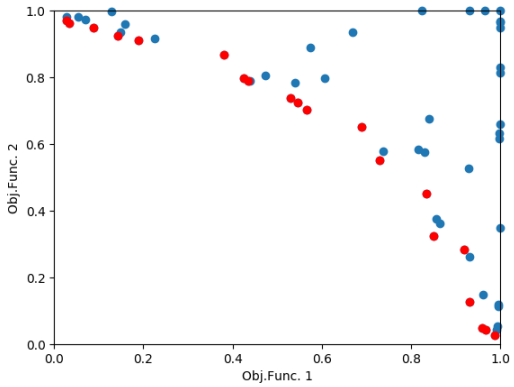
Nevertheless, a single run on each of two problems only provides some support to our initial conjecture that it is possible to guide optimization by directly modeling the hypervolume contribution without modeling the actual objective functions. However, a proper experimental analysis would be necessary to reach any definitive conclusions.




**(a)** Method 1a: $|ND| = 30$, $HV = 0.266164$




**(b)** Method 2a: $|ND| = 31$, $HV = 0.268667$




**(c)** ParEGO: $|ND| = 20$, $HV = 0.289163$

**Figure 20** Solutions, show in the objective space, produced by each approach when optimizing the Double Sphere problem after a maximum of 60 solution evaluations. Red dots indicate nondominated solutions, while blue dots are dominated ones.

### 4.5.4 A surrogate model based on binary $\epsilon$-metric

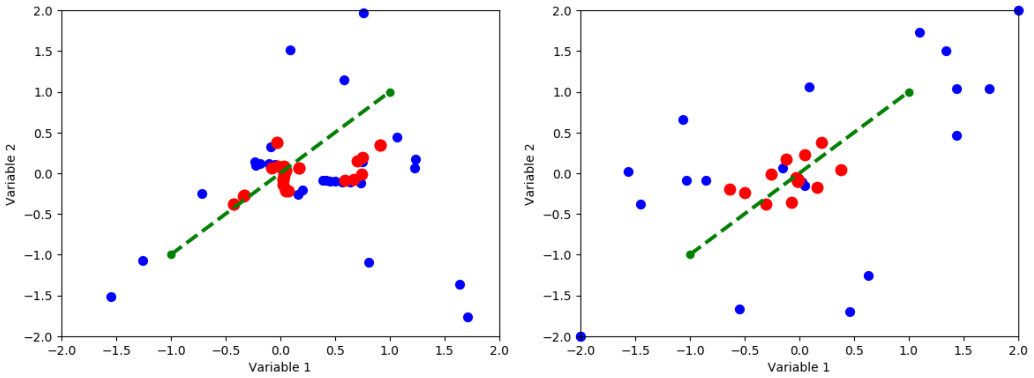
As shown above, trying to directly predict the hypervolume contribution requires the definition of alternative, but related metrics to assign a value to each point of our solution set $X_t$, since the actual $HVC$ value of those points would be zero. Instead of considering the hypervolume contribution, a different approach to multi-objective model-based optimization was discussed in our working group: The direct usage of the fitness function in IBEA [12] as the objective function. This fitness function is defined as

$$F(\vec{x}_1) = \sum_{\vec{x}_2 \in X_t \setminus \{\vec{x}_1\}} -e^{-I(\vec{x}_2, \vec{x}_1)/\kappa} \tag{23}$$

where $\kappa$ is a normalization parameter and the metric $I()$ above may be, for example, the additive binary $\epsilon$-metric:

$$I_{\epsilon+}(\vec{x}_2, \vec{x}_1) = \max_{i=1,\ldots,M} f_i(\vec{x}_2) - f_i(\vec{x}_1) \tag{24}$$

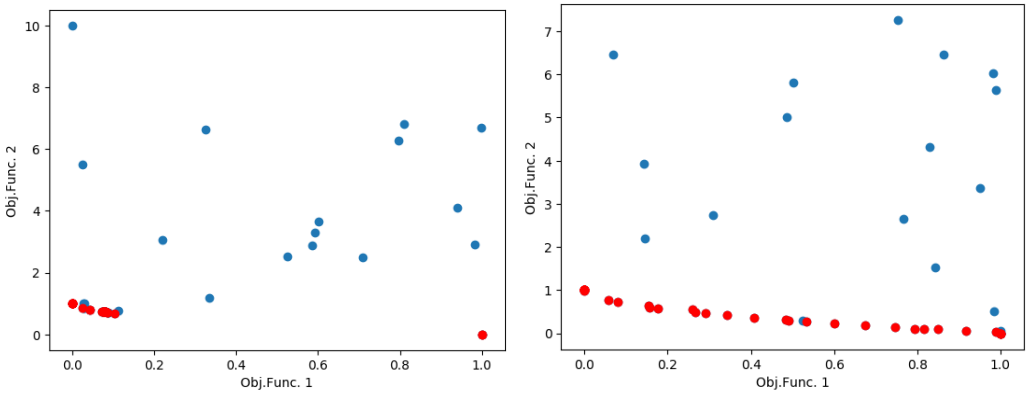
The benefit of the above fitness metric is that it naturally assigns a value to every point in our solution set, and those values will usually be different, except for specific solution



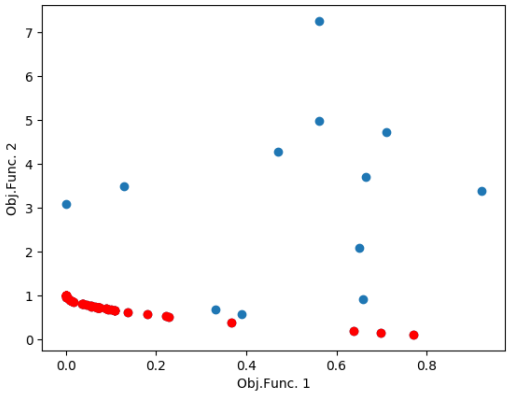

**(a)** Method 1a **(b)** Method 2a

**(c)** ParEGO

**Figure 21** Solutions, shown in the decision space, produced by each approach when optimizing the Double Sphere problem after a maximum of 60 solution evaluations. Red dots indicate nondominated solutions, while blue dots are dominated ones. The green dashed line corresponds to the optimal Pareto set.




**(a)** Method 1a: $HV = 120.297457$ **(b)** Method 1b: $HV = 120.631212$




**(c)** ParEGO: $HV = 119.385955$

**Figure 22** Solutions, show in the objective space, produced by each approach for the ZDT1 problem after 60 solution evaluations. Red dots indicate nondominated solutions, while blue dots are dominated ones.

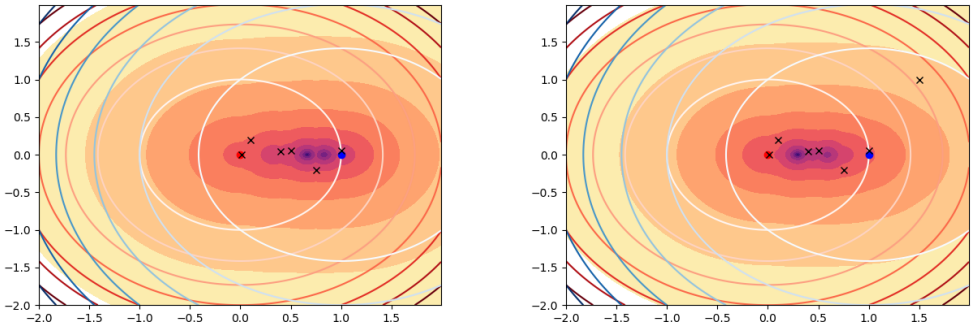


**Figure 23** Value of IBEA's fitness function $F()$ (using additive binary $\epsilon$-indicator) at each point of the decision space of $\mathbb{R}^2$ with respect to the solution set denoted by the symbols $\times$, when minimizing two Sphere functions, whose optimal solutions in $(0, 0)$ and $(1, 0)$ respectively correspond to the blue and red points. Darker colors indicate larger values of $F()$. The right plot shows the landscape of $F()$ after adding one (dominated) point to the solution set of the left plot.

sets that are unlikely to arise in real-world problems. In addition, IBEA has been show to perform consistently well (when properly tuned for the scenario at hand) in a large number of scenarios, often outperforming more recent and popular multi-objective evolutionary algorithms [3]. Thus, this fitness function is likely to produce a similarly well-performing Bayesian optimizer.

A quick numerical experiment, however, showed that the IBEA fitness function has the disadvantage of fundamentally changing its landscape after adding dominated points to the solution set into the history. Figure 23 shows the landscape of function $F(\vec{x})$ (Eq. 23) on a bi-dimensional decision space when optimizing two Sphere functions, with darker colors corresponding to higher values of $F()$. The optimal solutions of each Sphere function are shown as a red and a blue point, respectively, and contour lines denote the function value of each Sphere function. The current solution set $X_t$ is denoted by $\times$. The left plot shows the landscape of $F()$ with respect to five solutions in $X_t$. The right plot shows the landscape after adding an additional (dominated) solution to $X_t$ at the top right. The difference in colors between the two plots show that the landscape of the fitness function $F()$ has changed after adding this point, in particular, the peaks of the function have shifted towards the red point.

### 4.5.5 Conclusions

Many optimization problems have objective functions that are expensive to evaluate. In this case, meta-modeling allows predicting where to look for next solutions to be evaluated. The insights of newly evaluated solutions are then taken into account to update or refine the model for the problem at hand. Existing approaches either model the individual objective functions or a weighted aggregation thereof, however, we are not aware of any attempts at modeling directly the quality metrics that guide several multi-objective evolutionary algorithms.

In this report, we have discussed several ways to directly model the hypervolume contribution. Preliminary results on two problems suggest that this approach can guide a Bayesian multi-objective optimizer based on Gaussian Processes (GP), however, we also identified that the solutions generated have a low diversity and appear clustered in small regions of the decision and objective spaces. In addition, we also proposed how to model the fitness

function of IBEA, which is based on the binary $\epsilon$-indicator. This $\epsilon$-based fitness seems, in principle, easier to model directly than the hypervolume contribution, being able to directly provide a value for every point evaluated by the algorithm.

Further work is necessary to determine the advantages and disadvantages of the variants proposed here and empirically analyze their performance on multiple problems.

## References


1 Opto – a python framework for optimization. https://github.com/robertocalandra/opto
2 Aghamohammadi, N.R., Salomon, S., Yan, Y., Purshouse, R.C.: On the effect of scalarising norm choice in a ParEGO implementation. In: Trautmann, H., Rudolph, G., Klamroth, K., Schütze, O., Wiecek, M.M., Jin, Y., Grimme, C. (eds.) Evolutionary Multi-criterion Optimization, EMO 2017, pp. 1–15. Lecture Notes in Computer Science, Springer, Cham (2017)
3 Bezerra, L.C.T., López-Ibáñez, M., Stützle, T.: A large-scale experimental evaluation of high-performing multi- and many-objective evolutionary algorithms. Evolutionary Computation (2017), to appear
4 Emmerich, M.: Single-and multi-objective evolutionary design optimization assisted by gaussian random field metamodels. Dissertation, LS11, FB Informatik, Universität Dortmund, Germany (2005)
5 Emmerich, M.T.M., Yang, K., Deutz, A., Wang, H., Fonseca, C.M.: A multicriteria generalization of bayesian global optimization. In: Pardalos, P.M., Zhigljavsky, A., Žilinskas, J. (eds.) Advances in Stochastic and Deterministic Global Optimization, pp. 229–242. Springer International Publishing, Cham (2016)
6 Horn, D., Wagner, T., Biermann, D., Weihs, C., Bischl, B.: Model-based multi-objective optimization: taxonomy, multi-point proposal, toolbox and benchmark. In: Evolutionary Multi-Criterion Optimization (EMO 2015). pp. 64–78. Springer (2015)
7 Knowles, J.D.: ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. IEEE Transactions on Evolutionary Computation 10(1), 50–66 (2006)
8 Loshchilov, I., Schoenauer, M., Sebag, M.: A mono surrogate for multiobjective optimization. In: Genetic and Evolutionary Computation Conference (GECCO 2010). pp. 471–478. ACM (2010)
9 Ponweiser, W., Wagner, T., Biermann, D., Vincze, M.: Multiobjective optimization on a limited budget of evaluations using model-assisted $\mathcal{S}$-metric selection. In: International Conference on Parallel Problem Solving from Nature. pp. 784–794. Springer (2008)
10 Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. The MIT Press (2006)
11 Zitzler, E., Brockhoff, D., Thiele, L.: The hypervolume indicator revisited: On the design of Pareto-compliant indicators via weighted integration. In: Obayashi, S., et al. (eds.) Evolutionary Multi-criterion Optimization, EMO 2007, Lecture Notes in Computer Science, vol. 4403, pp. 862–876. Springer, Heidelberg, Germany (2007)
12 Zitzler, E., Künzli, S.: Indicator-based selection in multiobjective search. In: Yao, X., et al. (eds.) Proceedings of PPSN-VIII, Eigth International Conference on Parallel Problem Solving from Nature, Lecture Notes in Computer Science, vol. 3242, pp. 832–842. Springer, Heidelberg, Germany (2004)
13 Zitzler, E., Thiele, L., Deb, K.: Comparison of multiobjective evolutionary algorithms: Empirical results. Evolutionary Computation 8(2), 173–195 (2000)
14 Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Grunert da Fonseca, V.: Performance assessment of multiobjective optimizers: an analysis and review. IEEE Transactions on Evolutionary Computation 7(2), 117–132 (2003)

## 5 Topics of Interest for Participants for the Next Dagstuhl Seminar

During the summary session on Friday, all participants had an extensive discussion on the future challenges related to EMO and MCDM. This has lead to a plethora of suggestions for future seminar topics continuing the series. Photographs of topics of interest for participants for the next Dagstuhl seminar on EMO & MCDM are shown in Figure 24. The suggestions will be used by the organizers towards the proposal for a continuation of the series.

## 6 Changes in the Seminar Organization Body

### Joshua Knowles steps down as co-organizer

On behalf of all the participants of the seminar, KK, GR and MW would like to extend our warm thank you to Joshua Knowles for his contributions to this Dagstuhl seminar series on Multiobjective Optimization as he steps down from the role of co-organizer, which he has held for three terms of office. To our large regret, Joshua could not be in Dagstuhl during the seminar week. Nevertheless, he has played a leading role in shaping the topic, sharpening the research questions and setting us all up on an exciting journey to personalization. We are very thankful for his advice and activities in the preparation of this and the previous seminars. Thank you, Joshua!

### Welcome to Carlos Fonseca

We are very pleased that our esteemed colleague Carlos Fonseca has agreed to serve as co-organizer for future editions of this Dagstuhl seminar series on Multiobjective Optimization.

## 7 Seminar Schedule

### Monday, January 15, 2018

**08:45 – 10:30: Welcome Session**

- Welcome and Introduction
- Short presentation of all participants (2 minutes each!)
- Introduction to the topic of the seminar

**Coffee Break**

**11:00 – 12:00: Application Challenges**

- Karl Heinz Küfer: Industrial Applications of Multicriteria Decision Support Systems
- Georges Fadel: Culturally Tailored Multicriteria Product Design using Crowdsourcing

**Lunch**

**13:30 – 14:30: Personalization in Model Building, Approximation, and Representation**

- Kalyanmoy Deb: Metamodeling Approaches for Multiobjective Optimization
- Serpil Sayin: Representations: Do they have Potential for Customer Choice?

**Coffee Break**

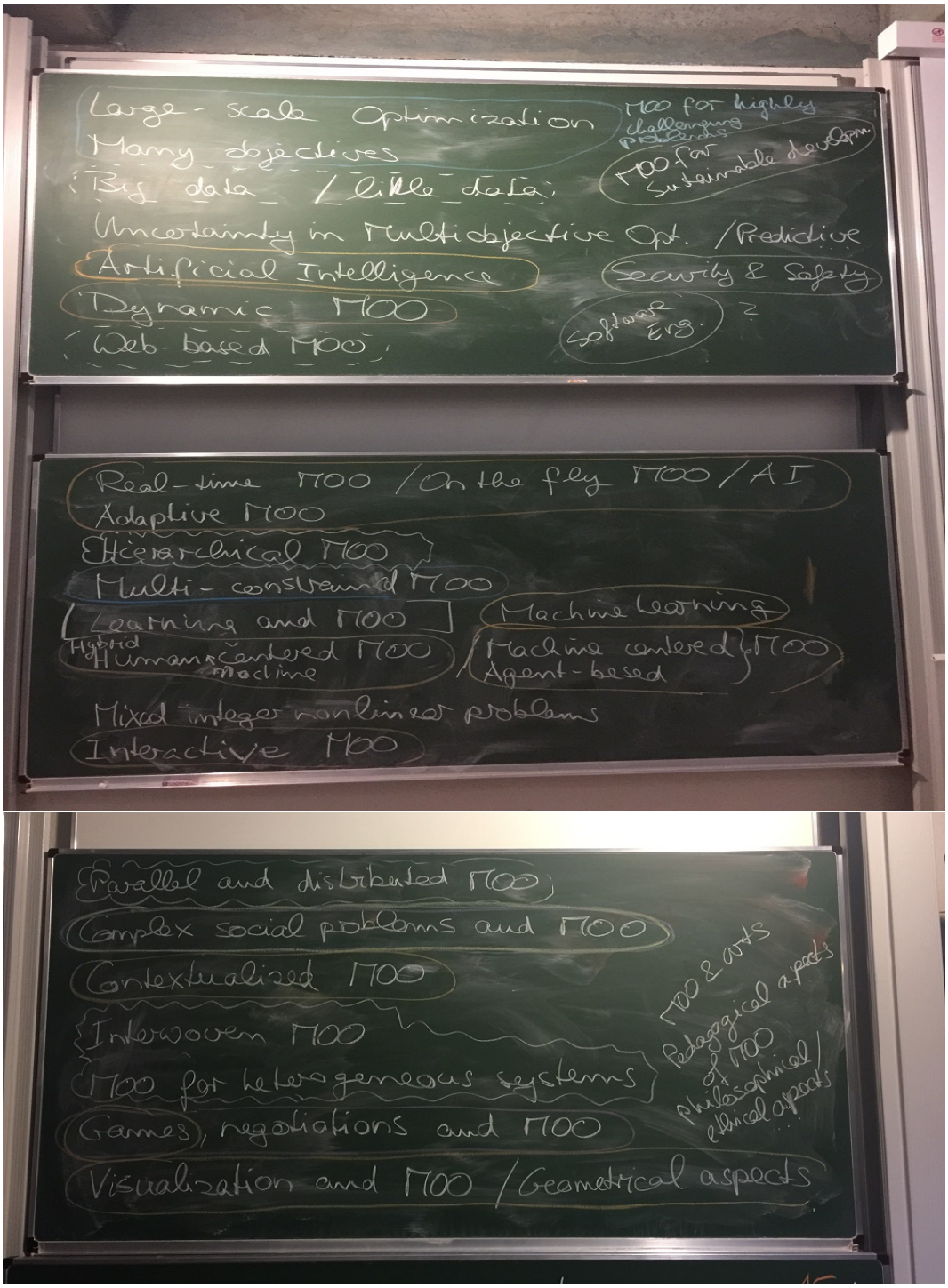



**Figure 24** Topics of interest for participants for the next Dagstuhl seminar.

**15:00 – 15:30: Personalization and Preference Modelling**
- Robin Purshouse: Modelling Complex Networks of Decision Makers: An Analytical Sociology Perspective

**15:30 – 16:00: Personalization in Algorithm Design and Efficiency**
- Manuel López-Ibánez: Data-Driven Automatic Design of Multi-Objective Optimizers

**Break**

**16:15 – 18:00: Group Discussion about Hot Topics and Working Groups**

## Tuesday, January 16, 2018

**09:00 – 10:00: Decision Analytics and Consensus** Chair: Salvatore Greco
- Michael Emmerich: Maximizing the Probability of Consensus in Group Decision Making
- Kaisa Miettinen: Decision Analytics with Multiobjective Optimization and a Case in Inventory Management

**Coffee Break**

**10:30 – 12:00: Working Groups**

**Lunch**

**13:30 – 14:30: Personalization and Learning** Chair: Jussi Hakanen
- Jürgen Branke: Active Learning for Mapping Advertisements to Customers
- Roman Slowinski: The NEMO framework for EMO: Learning value functions from pairwise comparisons

**Coffee Break**

**15:00 – 17:00: Working Groups**

**17:00 – 18:00: Reports from Working Groups**
- 6 minutes / 3 slides per working group
- General discussion and working group adaptations

## Wednesday, January 17, 2018

**09:00 – 10:00: Metamodelling and Knowledge Extraction** Chair: Carlos Fonseca
- Mickaël Binois: Uncertainty Quantification on Pareto Fronts
- Abhinav Gaur: Unveiling Invariant Rules from Non-Dominated Solutions for Knowledge Discovery and Faster Convergence

**10:00: Announcements**

**Coffee Break**

**10:30 – 12:00: Working Groups**

**Lunch**

**14:00: Group Foto (Outside)**

**14:05 – 16:00: Hiking Trip**

**16:30 – 18:00: Reports from Working Groups**
- 15 minutes / 5 slides per working group

## Thursday, January 18, 2018

**9:00 – 10:00: Data Structures** Chair: Christoph Lofi
- José Rui Figueira: Compressed Data Structures for Bi-Objective $\{0,1\}$-Knapsack Problems
- Andrzej Jaszkiewicz: Recent Algorithmic Progress in Multiobjective (Combinatorial) Optimization

**Coffee Break**

**10:30 – 12:00: Working Groups**

**Lunch**

**13:30 – 15:30: Working Groups**

**Coffee Break**

**16:00 – 17:00: Working Groups**

**17:00 – 18:00: Continuing the Dagstuhl Seminar Series**

**20:00: Wine & Cheese Party** (Music Room)

## Friday, January 19, 2018

**9:00 – 11:00: Presentation of Working Group Results**

**Coffee Break**

**11:30 – 12:00: Summary, Feedback, and Next Steps**

**Lunch & Goodbye**

## Participants


- Richard Allmendinger
University of Manchester, GB
- Mickaël Binois
Argonne National Laboratory – Lemont, US
- Jürgen Branke
University of Warwick, GB
- Dimo Brockhoff
INRIA Saclay – Palaiseau, FR
- Roberto Calandra
University of California – Berkeley, US
- Carlos A. Coello Coello
CINVESTAV – Mexico, MX
- Kerstin Dächert
Universität Wuppertal, DE
- Kalyanmoy Deb
Michigan State University, US
- Matthias Ehrgott
Lancaster University Management School, GB
- Gabriele Eichfelder
TU Ilmenau, DE
- Michael Emmerich
Leiden University, NL
- Alexander Engau
Lancaster University Management School, GB
- Georges Fadel
Clemson University – Clemson, US
- José Rui Figueira
IST – Lisbon, PT
- Carlos M. Fonseca
University of Coimbra, PT
- Abhinav Gaur
Michigan State University – East Lansing, US
- Salvatore Greco
University of Catania, IT
- Jussi Hakanen
University of Jyväskylä, FI
- Johannes Jahn
Universität Erlangen-Nürnberg, DE
- Andrzej Jaszkiewicz
Poznan University of Technology, PL
- Milosz Kadzinski
Poznan University of Technology, PL
- Kathrin Klamroth
Universität Wuppertal, DE
- Karl Heinz Küfer
Fraunhofer ITWM – Kaiserslautern, DE
- Christoph Lofi
TU Delft, NL
- Manuel López-Ibánez
University of Manchester, GB
- Kaisa Miettinen
University of Jyväskylä, FI
- Sanaz Mostaghim
Universität Magdeburg, DE
- Boris Naujoks
TH Köln, DE
- Frank Neumann
University of Adelaide, AU
- Luís Paquete
University of Coimbra, PT
- Robin Purshouse
University of Sheffield, GB
- Patrick M. Reed
Cornell University, US
- Günter Rudolph
TU Dortmund, DE
- Stefan Ruzika
TU Kaiserslautern, DE
- Serpil Sayin
Koc University – Istanbul, TR
- Pradyumn Kumar Shukla
KIT – Karlsruher Institut für Technologie, DE
- Roman Slowinski
Poznan University of Technology, PL
- Ralph E. Steuer
University of Georgia, US
- Theodor J. Stewart
University of Cape Town, ZA
- Michael Stiglmayr
Universität Wuppertal, DE
- Lothar Thiele
ETH Zürich, CH
- Selvakumar Ulaganathan
Noesis Solutions – Leuven, BE
- Daniel Vanderpooten
University Paris-Dauphine, FR
- Margaret M. Wiecek
Clemson University, US


Report from Dagstuhl Seminar 18041

# Foundations of Data Visualization

**Edited by**

**Helwig Hauser[1], Penny Rheingans[2], and Gerik Scheuermann[3]**

1 **University of Bergen, NO,** helwig.hauser@uib.no
2 **University of Maryland, Baltimore County, US,** rheingan@umbc.edu
3 **Universität Leipzig, DE,** scheuermann@informatik.uni-leipzig.de

## Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 18041 "Foundations of Data Visualization". It includes a discussion of the motivation and overall organization, an abstract from each of the participants, and a report about each of the working groups.

**Seminar** January 21–26, 2018 – https://www.dagstuhl.de/18041
**2012 ACM Subject Classification** Human-centered computing → Empirical studies in visualization, Human-centered computing → Visualization design and evaluation methods, Human-centered computing → Visualization theory, concepts and paradigms
**Keywords and phrases** Foundations, Interdisciplinary Cooperation, Theory, Visualization
**Digital Object Identifier** 10.4230/DagRep.8.1.100

## 1 Executive Summary

*Penny Rheingans (University of Maryland, Baltimore County, US)*



Data Visualization is the transformation of data, derived from observation or simulation, and models into interactive images. It has become an indispensable part of the knowledge discovery process in many fields of contemporary endeavor. Since its inception about three decades ago, the techniques of data visualization have aided scientists, engineers, medical practitioners, analysts, and others in the study of a wide variety of data, including numerical simulation based on high-performance computing, measured data from modern scanners (CT, MR, seismic imaging, satellite imaging), and survey and sampled data, and metadata about data confidence or provenance. One of the powerful strengths of data visualization is the effective and efficient utilization of the broad bandwidth of the human sensory system in interpreting and steering complex processes involving spatiotemporal data across a diverse set of application disciplines. Since vision dominates our sensory input, strong efforts have been made to bring the mathematical abstraction and modeling to our eyes through the mediation of computer graphics. The interplay between these multidisciplinary foundations of visualization and currently emerging, new research challenges in data visualization constitute the basis of this seminar.

The rapid advances in data visualization have resulted in a large collection of visual designs, algorithms, software tools, and development kits. There is also a substantial body of work on mathematical approaches in visualizations such as topological methods, feature extraction approaches, and information theoretical considerations. However, a unified description of theoretical and perceptual aspects of visualization would allow visualization practitioners to derive even better solutions using a sound theoretical basis. There are promising ideas



but they need further discussion. Currently, we employ user studies to decide if a visual design is more effective, but a comprehensive theory would allow visualization researchers to answer why one visual design is more effective than another and how the visual design can be optimized. Furthermore, we usually have an understanding of the role of a specific visualization in a specific analytic workflow, but we would like to formalize the general role of visualization in the analytic workflow. This would also allow for more quantitative measures of visualization quality. In addition, the community needs a deeper, general understanding of the most informative way to conduct perceptual and usability studies involving domain experts.

For this seminar, we chose to take a focused consideration of the foundations of visualization in order to establish an integrated discussion on the fundamental understanding and generic methodologies of data visualization, including theories, models and workflows of data visualization, evaluation metrics, and perceptual and usability studies. We included experts from all areas of visualization such as scientific visualization, information visualization, and visual analytics to allow for an in-depth discussion of our shared research foundations based on a broad expertise.

With the experience of delivering technical advances over the past three decades, it is timely for the visualization community to address these fundamental questions with a concerted effort. Such an effort will be critical to the long-term development of the subject, especially in building a theoretical foundation for the subject. The community needs to develop suitable models for the whole visualization process from cleaning and filtering the data, analysis processing, mapping to graphical scenes, to the interpretation by the human visual system. While there are some methods of evaluation based on user studies and findings in applications, a complete theoretical foundation for evaluations is missing. Modern visualization includes advanced numerical and combinatorial data processing, so the correctness of this processing including a critical look at its assumptions with respect to the application at hand is needed. Only then, visualization can establish strong correlations between visualization algorithms and questions in the application domains. In addition, uncertainty has received attention from the visualization community in recent years, but a full analysis of uncertainty at all stages of the established visualization pipeline is still not available. Theoretical foundations of uncertainty in visualization need to look at uncertainty in the data, errors due to numerical processing, errors due to visual depiction and, finally, uncertainty in the results based on human misinterpretation of interactive visual depictions.

This workshop addressed five important topics:

Theory of overall visualization process. A theory of the whole visualization process needs to cover all parts of the visualization pipeline and should be applicable to broad classes of application domains. Of course, it is the ultimate foundation, but there are a few formulation attempts and the seminar discussed them. Such a theory should allow to find optimal visualizations and to quantify the value of visualizations. In addition, it is strongly believed by most experts that such a theory needs to cover the challenge of uncertainty in the data, the processing including visual mapping and potential misinterpretation by human observers.

Foundations of evaluation. Evaluation allows designers and analysts to select visualization approaches from among different options for a specific problem. One evaluation method is a user study, usually with a larger group of subjects. Here, it is often a challenge that there is only a very small set of experts available that understand the scientific questions behind the data. Guidelines for user study design in these situations are necessary. In addition, evaluation needs to look at limits of the human visual system. In advanced analytic applications, it is also very important to study the relation between user interest

and visualization. There are many open questions in this area that will be discussed in the seminar.

Collaboration with domain experts. Many visualizations address questions and needs from expert researchers, engineers, analysts, or decision makers. Therefore, visualization nearly always involves people outside the visualization community. The seminar included some representatives from large applied research centers so that the discussion about relations between visual data analysis and application semantics was not carried out without domain experts. These participants also commented on methodologies for defining domain requirements and realistic roles of application researchers in evaluation.

Visualization for broad audiences. Visualizations developed for broad audiences involve context and constraints different from those developed for expert domain collaborators. Such visualizations include those for personal information, school use, science centers and other public settings, and communication with a broad general public. Issues with developing visualizations for broad audiences include a higher need for intuitive metaphors and conventions, a larger imperative for drawing participants into interaction, and more requirements for robust interfaces and systems.

Mathematical foundations of visual data analysis. There is a rich tradition of mathematical/computational methods used in visualization, such as topological approaches, mathematical descriptions of feature extraction, numerical sampling and reconstruction methods, integration, differential operators, filtering, dimension reduction, and applications of information theory. In addition, we have seen promising attempts to incorporate uncertainty in these mathematical approaches. While all these methods have a solid mathematical foundation, a careful look at the relation between theories in applications and these mathematical approaches in visual data analysis was taken in this seminar.

The format of the seminar incorporated several elements: overview talks on each topic, clusters of short talks on a single topic followed by a joint panel discussion, and breakout groups on each of the five topics. Unlike the typical arrangement, all presentations in each session were given in sequence without a short Q&A session at the end of each talk. Instead, all speakers of a session were invited to sit on the stage after the presentations, and answer questions in a manner similar to panel discussions. This format successfully brought senior and junior researchers onto the same platform, and enabled researchers to seek a generic and deep understanding through their questions and answers. It also stimulated very long, intense, and fruitful discussions that were embraced by all participants. The breakout groups focused on the general themes and are reported in a later section.

## 2 Table of Contents

# 3 Overview of Talks

## 3.1 Towards a Theory for Massive, Multidimensional Data Analysis and Visualization

*James Ahrens (Los Alamos National Lab., US)*



Sensors and simulations are producing massive amounts of multidimensional data on the order of 10$\hat{1}$2-10$\hat{1}$8 bytes that need to be visualized and understood. The human visual system can process data on the scale of the order of 10$\hat{6}$. Therefore some type of data reduction or sampling is required to produce a visualization.

In this talk, I focus on in situ visualization, visualizing data while it is being generated by a simulation on a supercomputer. Three in situ approaches that use sampling are presented. The first approach, which we refer to as Cinema, conceptually visualizes all results needed while simulation data is in memory for later exploration. Results are generated via rendering a complete Cartesian project of all interesting operators, parameters and camera positions. Results are selected via a set of sliders for the parameters. The second approach extends Cinema to sparse experimental data using parallel coordinates to identify and select the sparse entries. The third approach, proposes the use of a sample-based data representation as a common representation for all data. Each visualization operator inputs and outputs samples. A pipeline-based composition of these operators reduces data to the target size.

## 3.2 Who Are We (in a Collaboration)?

*Johanna Beyer (Harvard University - Cambridge, US)*



Visualization researchers and practitioners face many challenges when collaborating with domain experts. In particular, the different roles of visualization researchers as compared to visualization engineers or practitioners have a huge influence on the goals and measures for success for a collaboration. While visualization researchers should focus on novel algorithms, tools, and ultimately publications in visualization-related fields, the main focus of visualization engineers generally lies in creating usable software tools that are used beyond the initial prototype stage. Therefore, these different roles should be explicitly addressed at the beginning of a collaboration, to avoid common pitfalls and differing expectations between collaborating visualization and domain experts.

## 3.3 A Model of Spatial Directness in Interactive Visualization

*Stefan Bruckner (University of Bergen, NO)*



The ability to interactively explore a visual representation is a core aspect of all visualization systems. The term “directness”, as in “direct manipulation”, is commonly used to discuss

properties of interaction techniques in the context of visualization. Unfortunately, the terms referring to the directness of spatial interaction are largely used by intuition and without a clear definition. In this talk, I introduce a model of directness in interactive visualization that characterizes it as an emerging property of the involved mapping processes, from the data space to the perception and cognition of the user. Based on such a formulation, we can further proceed to quantify the different dimensions of directness, leading us to an approach that forms the basis of formulating testable predictions for visualizations that may ultimately allow us to perform in-silico user studies and even allow the synthesis of novel visualization methods based on different objective functions.

## 3.4 Color, Math, and Visualization

*Roxana Bujack (Los Alamos National Laboratory, US)*



Perceptual scientists' experiments indicate that human color perception is non-Euclidean, which induces new challenges on colormap design. How can we generalize methods for the evaluation, optimization, generalization and interpolation of colormaps?

## 3.5 Visualization of Climate Projections for Communication to the Public

*Michael Böttinger (DKRZ Hamburg, DE)*



Increasing public attention to climate and climate change triggers a demand by the media, policy makers and the general public for meaningful visualizations showing key outcomes of future climate simulations. At the German Climate Computing Center (DKRZ), visualizations of IPCC simulations have regularly been produced for more than 20 years. One of the keys for successful visualization is simplicity. However, to help recipients of these visualizations in identifying the main outcomes, accompanying annotation in the form of text or narration proved to be useful.

In this talk, several examples of successful visualizations are discussed which had been adopted by various media. The latest example refers to one of the key visualizations of the IPCC AR5 summary for policymakers that shows the temperature change of the CMIP5 multi model ensemble with two levels of robustness overlaid by stippling and hatching. We present an alternative, simplified and animated version for the same data set that draws the attention of the viewer to robust areas by dehighlighting non-robust areas. In this way, the viewer's focus is guided to the trustworty part of the data.

## 3.6 My Math Keeps Breaking!

*Hamish Carr (University of Leeds, GB)*



Visualization relies heavily on mathematics, both because the input data is often defined mathematically, and because the mathematics is the tool that we use to describe the stages of data processing in our data pipelines. However, the complexity of our pipelines and the nature of the mathematical computations we perform causes increasing problems in our mathematics. One way this occurs is that different stages in the computation are often handled by different people, with different mathematical assumptions. In the worst case, their mathematical assumptions are irreconcilable, but even when they are formally reconcilable, their cumulative effect is to make the overall computation unreliable. Moreover, much of the mathematics we use has formal assumptions that are computationally difficult or impossible to guarantee, leading to the need for new mathematics.

## 3.7 Empirical Studies in Visualization

*Min Chen (University of Oxford, GB)*



In the field of visualization, empirical studies are typically conducted under the major scope of "Evaluation". The emphases have typically been placed on "testing" some visual designs or visualization systems as part of a software engineering workflow. While empirical studies can and should support "evaluation" in visualization, there have not been enough emphases given to the more ambitious goal of empirical studies, that is, to make new discoveries about how and why visualization works in some conditions and not in others, and to inform and verify proposed theories advances.

Most of us agree that in some circumstances, visualization is more effective and/or efficient than viewing data in numerical, textual, or tabular forms, and than being simply informed by a computer about the decision. When visualization works in these circumstances, there must be some merits in perception and cognition. Hence any causal factors that make visualization work may potentially be the causal factors that make perception and cognition work. Therefore, visualization researchers are in the right place at the right time to look for these causal factors. For example, can the cost-benefit metric proposed by Chen and Golan also be the fitness function for the development or evaluation of some perceptual and cognitive capabilities (e.g., visual search, selective attention, gestalt grouping, heuristics, and memory)?

## 3.8 Topological data analysis and topology-based visualization

*Leila De Floriani (University of Maryland - College Park, US)*



The talk deals with common topological and algorithmic tools to two Topological Data Analysis and topology-based visualization. Specifically, it focuses on an algebraic topology tool, Discrete Morse Theory (DMT), applied in both disciplines, and its relation with persistent homology. New developments in dealing with multivariate data which led to multi-parameter persistent homology in TDA are discussed as well a new approach for computing multi-parameter persistent homology and its possible applications to critical feature extraction in multifield data analysis and visualization.

## 3.9 Fundamental Mathematics in Visualization

*Christoph Garth (TU Kaiserslautern, DE)*



Mathematical concepts, methods, and tools have played a key role in the development of a large variety of visualization techniques. This raises the question, which mathematical techniques should visualization researchers be familiar with. In my talk, I will report on an informal survey of the mathematical underpinnings of the past decade of visualization research, and examine its implications for the education of students.

## 3.10 Adjust, Just Adjust

*Eduard Gröller (TU Wien, AT)*



Developing visualizations for broad audiences requires glanceable graphics and graspable interactions. This talk will concentrate on interaction facilitation through automatic adjustments. The first example illustrates an automatic color scale adjustment in a biomolecular setting to accommodate contradicting and overlapping color schemes across scales. The second example discusses output-sensitive interaction to make changes in the input proportional to changes in the output, or to visually indicate the sensitivity of input changes with respect to output changes. The third example deals with visualization of 4D ultrasound data, which is targeted to a broad audience in prenatal imaging and diagnosis. Lessons learned during this project are presented. The talk makes a case for automatically reducing interaction complexity in visualizations for broad audiences.

## 3.11 Visualize Insight??

*Hans Hagen (TU Kaiserslautern, DE)*



One purpose of data visualization is to help the viewer to obtain insight. But how does insight emerge from data? Is insight part of the visualization? Can we somehow characterize the insights to be found in a visualization?

## 3.12 Effective Collaboration with Domain Experts: FluoRender

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



Effective collaboration with domain experts requires knowledge, joint interest, and coordination to achieve joint scientific goals. FluoRender is an example of close collaborations with biologists and visualization experts that resulted in a widely used visualization tool that has contributed results to the visualization community and enabled scientific results in the biology field. There are several lessons that can be learned from this collaboration. First, communication is key and a common language and vocabulary is fundamental. Both parties, visualization and domain experts, should accomplish scientific contributions in their respected fields. Close collaboration requires detailed application knowledge by the visualization research and visualization knowledge by the domain expert. It is important not to ask the domain expert what problems need solving or which features are required. It is better to understand the domain scientist's workflow by spending time in their research laboratory, work closely with them and do not limit interaction to meetings and discussions. Lastly, it is important to be creative, have fun, and collaborate.

## 3.13 About the scales and limits of visualization

*Helwig Hauser (University of Bergen, NO)*



Discussing what works in visualization, and what not, can be done in terms of several principal aspects of influence. At the side of the user, perceptual and cognitive aspects of the non-uniform human visual system are important, enabling (and also limiting) visualization. Then, of course, the extent of the data has a major influence and there is a certain range of extent that lends itself to visualization solutions. Similarily, the richness of the data, for example, in terms of multivariate data is criticial. Thirdly, in this respect, the dimensionality of the data leads to major differences – a few dozens of dimensions are very different from hundreds! Last, but not least, a more technical perspective is impoortant: good hardward and good software. All in all, these aspects possibly form a space for visualization solutions.

## 3.14 Benefits of and Questions to a Theory of Visualization

*Hans-Christian Hege (Zuse Institute Berlin, DE)*



In this talk, "theory" does mean no single universal theory (which might not be achievable) but a bundle of theories. What would the benefits of such a theory building be and which fundamental questions should it help to answer?

The practical benefits are obvious: Qualitative, conceptual models could help us to describe, understand and reason about visualization processes and provide hints, which visual representations, analysis actions, and work flows are more efficient than others. Quantitative models would help us, to make predictions about quantitative dependencies in visualization processes and help us to optimize mathematically components of visualization processes.

Beside that there are strategic benefits; in particular, a common core theory would be an effective countermeasure to the danger of fragmentation of data visualization. It would also increase its survival capability in the landscape of competing disciplines.

The list of fundamental questions to be answered, is long. Here are some, commented in the talk: What is visualization and what is it for? What is information? What is the solution of the data–information–knowledge conundrum? What are the elementary knowledge units? How can prior knowledge be captured in detail? What are the elementary acts of reasoning, given new external information? What are the elementary acts in which humans increase their knowledge using external (nonvisual/visual) information? What leads to the emergent phenomenon of a eureka moment? How are external and mental images related? What is the role of mental images in reasoning? What are the limits of visualization? If we have (better) answers to such questions: How can the theories be made operational and how can they be practically utilized?

Almost all these fundamental questions can be answered only in collaboration with other sciences, especially cognitive sciences. That will not happen, if it is not initiated by us.

## 3.15 Theory of Visualization and Domain Experts

*Mario Hlawitschka (Hochschule Leipzig, DE)*



Working with "domain scientists" may be challenging, especially when they come from different domains and a scientific language barrier must be lowered. On the first day of setting up a project, the group should be aware of what they can expect from their partners. Even in the field of visualization, the definition of visualization is rather vague and the first step of a fruitful collaboration is to explain the potentials of visualization. A "theory of visualization" could aid in finding good definitions of the process of visualization and its potentials. Guidelines should be derived from that, which should be used by domain experts as well as visualization designers and researchers. An example is "information theory" where a profound basis has been set in a very specific topic, which is now used in a much broader way. Such building blocks, both algorithmically but also as parts of a theory (or theories) of visualization lay a foundation and correctly applied, may lower that entry barrier. Ultimately, this may lead to one or many "theories of visualization" that may be a foundation to impact many other fields of research.

## 3.16 What I am thinking about when I am biking to work: Spaces – mappings – projections

*Ingrid Hotz (Linköping University, SE)*



This talk reasons about scientific knowledge discovery process as a sequence of mappings from and to various spaces. Spaces involved in such a pipeline could be defined by models, data, application areas, or humans exhibiting a certain experience. The mapping between these spaces can be of different nature preserving the dimension or reducing the dimensions describing projections. A carful design of these spaces, their parametrization and the mappings between is essential for the success of the process. Within the visualization pipeline one can exemplarily consider the data space representing the data according to some model space, the space spanned by relevant questions in one application, and a space used by the visualization.

## 3.17 Pathways for Theoretical Advances in Visualization

*Christopher R. Johnson (University of Utah - Salt Lake City, US)*



In my 2004 article, Top Scientific Visualization Research Problems, I proposed creating a Theory of Visualization as a top research problem. Since 2004, there has been some progress in theoretical aspects of visualization, but much more needs to be done in this area. In 2017, Min Chen lead a co-author team of M. Chen, G. Grinstein, myself, J. Kennedy, and M. Tory who proposed Pathways for Theoretical Advances in Visualization [1]. We hope that many visualization researchers will contribute to this foundational area within visualization.

### References

1 Chen, M., Grinstein, G., Johnson, C. R., Kennedy, J., and Tory, M. Pathways for theoretical advances in visualization. IEEE Computer Graphics and Applications, 37(4):103–112, 2017.

## 3.18 Empirical Studies with Domain Experts

*Alark Joshi (University of San Francisco, US)*



Tool adoption by domain users is a strong measure of success when working with domain experts. Working with domain experts requires deep, longer conversations that go from learning each others language to working closely on prototypes to help solve their problem. When working with atmospheric physicists, we developed a system to predict hurricane dissipation and even though we conducted formative and summative evaluations, it was eventually not adopted for regular use. In our collaboration with neurosurgeons, we developed a tool that works with an image-guided navigation system. We conducted various empirical studies to evaluate the use of a novel interaction technique for multimodal visualization,

applying existing visualization techniques for vascular visualization, and so on. Empirical studies can truly help you learn about specific aspects in your system/technique. I believe that empirical studies should not be an afterthought and working with human factors experts can help us design better studies to learn from them.

## 3.19 Making sense of Math in Vis

*Gordon Kindlmann (University of Chicago, US)*



Visualization research sometimes has a complicated relationship to mathematics. Many accounts of data visualization do not include a presentation or discussion of the underlying mathematics employed. When there is math, it can come myriad forms. The types of mathematics used for one type of research may or may not be similar to those for other research: the linear algebra for tensor visualization is distinct from the statistics used to measure the results of user studies. This talk attempts to locate the places *in* visualization where math arises, as well as outlining some recent work on the math *of* visualization. The necessity of math in visualization will likely remain an ongoing topic of consideration and debate.

## 3.20 Data-driven Storytelling at NASA

*Helen-Nicole Kostis (USRA/GESTAR SVS NASA/GSFC, US)*



In this talk, I will provide an overview of the storytelling efforts at NASA Goddard Space Flight Center. The goal of the Scientific Visualization Studio (SVS) is to promote greater understanding of NASA science programs through visualization. The products of the visualization efforts are data-driven high quality computer graphics animations that are developed and produced in collaboration between producers, science writers, visualization experts and scientists. NASA's heartbeat are the scientific results and engineering accomplishments. Through the years, data-driven visualizations from the Scientific Visualization Studio have clearly become a critical component on leading outreach, education and science communication efforts.

## 3.21 Collaboration with the Domain Experts - molecular visualization

*Barbora Kozlíková (Masaryk University - Brno, CZ)*



Understanding the structure and behavior of protein molecules is crucial in many biological and biochemical, such as drug design and protein engineering. This process requires studying the proteins from many aspects, including their constitution, physico-chemical properties,

temporal behavior, or interactions with other molecules. Observing that by traditional approaches, i.e., animation of the 3D structural model, is not feasible anymore, due to the amount of data to be processed. Therefore, specialized visualization techniques have to be involved into the exploration process. The talk covers short introduction to the domain problem and then focuses mainly on the experience in collaboration with the experts and lessons learned.

## 3.22 Accidental Broad Audiences in Virtual Reality Visualization

*David H. Laidlaw (Brown University - Providence, US)*



Over five minutes I will share some of the lessons learned showing several thousand audience members our large-scale virtual reality display and, within it, several scientific and academic applications we have developed. In particular, the short, pithy messages that are appropratie for broad audiences contrast with the more exploratory or formative activities that occur with our scientific research tool development. This has implications on the design of tools and how they are presented and used.

## 3.23 Foundations of visualization - Where we stand and where to go

*Heike Leitte (TU Kaiserslautern, DE)*



A theory is a set of scientifically founded statements used to describe a part of the world and make predictions about it. In the visualization field a number of such theories have been published over the last decades that help understand different aspects of the visualization process. Their validity, interrelationships, and impact have been discussed in a number of panels, but summarizing papers giving an overview over theories in and of visualization are scarce. Hence, it is time to join forces and structure the presented ideas, identify shortcomings, and think about future directions.

## 3.24 Empirical Studies on Human-in-the-Loop

*Ross Maciejewski (Arizona State University - Tempe, US)*



Currently, a large variety of empirical studies in information visualization have provided insights into how people perceive information, what the just noticeable differences are, response times, etc. However, less work has focused on understanding the use of knowledge being generated. This talk discusses issues in knowledge generation, open challenges, and the notion of algorithmic aversion and its potential relationship to visualization.

## 3.25 Activity-Centered Domain Characterization

*Georgeta Elisabeta Marai (University of Illinois - Chicago, US)*



Domain characterization is the first stage of the visualization process. Activity Theory helps lay an activity-centered foundation for this stage. In a departure from existing visualization models, this approach assigns value to a visualization based on user activities; ranks user tasks above user data; partitions requirements into activity-related capabilities and nonfunctional characteristics and constraints; and explicitly incorporates user workflows into the requirements process. A quantitative evaluation supports the merits of the activity-centered model and leads to several questions regarding the sparsity of the vis theoretical landscape, and about the evaluation models we use for theories.

## 3.26 Empirical Studies in Visualization

*Kresimir Matkovic (VRVis - Wien, AT)*



Empirical studies represents a well-established research field. Visualization researchers are often required to evaluate their research results. Empirical studies represent a possibility of evaluation. However, they are particularly suitable for well-defined tasks which can be easily quantizable (how long does it take to do something, what is in front what is behind, etc.). Such low level tasks are useful in visualization, but typical tasks are mostly more complex. How much knowledge is gained, what insights are gained, etc. Providing quantitative measures for such questions is not easy. This is why some of the visualization researches are not so enthusiastic with evaluation. A possible solution is to identify basic tasks and to test it by means of an empirical study. Another requirement which is often posed is to evaluate a specific technique developed for a specific domain in a user study and or to generalize it. Both requirements are not easy to fulfill. The experts are rare, so we cannot find enough of them for a proper user study. If we generalize it, we need a lot of users again. This might require additional resources (time) which are not always available. Finally, we often base evaluation on tasks abstraction. The data and user abstraction is usually neglected. Further, the tasks are rarely compared with similar tasks from peers' research. We argue, it is necessary to base the evaluation on abstraction of tasks, data, and users. Having a list of tasks, data and users with corresponding solution would be a valuable contribution to the visualization community.

## 3.27 Theory of Visualization Process: Survey? Overview? Challenges and Opportunities?

*Silvia Miksch (TU Wien, AT)*



Is there any (unified) theory of the visualization process, which is "somehow" accepted by the visualization communities? In this talk I present some definitions of visualization, with particular focus on the process characteristics, and various models in visualization to identify possible challenges and opportunities. Definitions and models on various levels of abstraction, functionality, and complexity exits, but no real unified ones. I propose a conceptual model of knowledge-assisted visual analytics incorporating the role of explicit knowledge as well as characterizing guidance in visual analytics.

### References

1 Federico, P., M. Wagner, A. Rind, A. Amor-Amorós, S. Miksch, and W. Aigner. The Role of Explicit Knowledge: A Conceptual Model of Knowledge-Assisted Visual Analytics. Proceedings of the IEEE Conference on Visual Analytics Science and Technology (IEEE VAST 2017). (http://www.cvast.tuwien.ac.at/node/785)

2 Ceneda, D., T. Gschwandtner, T. May, S. Miksch, H.-J. Schulz, M. Streit, and C. Tominski. Characterizing Guidance in Visual Analytics. IEEE Transactions on Visualization and Computer Graphics 23(1):111–120, 2017. (http://www.cvast.tuwien.ac.at/node/765)

## 3.28 Bridging the gap between domain experts and data analysts

*Daniela Oelke (Siemens AG - München, DE)*



An analyst needs an analysis question to work with. This question is provided by the domain expert, but has to be translated into a (more general) data analysis question before a data analyst can work with it. This process requires the data analyst and the domain expert to work closely together and can be challenging.

In my talk I presented experiences from industry of what has proven useful to bridge this gap including educating the domain expert, interviewing domain experts in the right way, and using visual analytics to facilitate the communication.

Furthermore, I pointed out that in order to have an impact, additional stakeholders or domain experts have to be included in the process in a company such as customers, sales representatives, management, etc. I reported on an experimental project in which all these stakesholders were working together for a week combining methods of visual data analytics and business innovation to come up with ideas for novel business opportunities.

## 3.29 I work with Experts

*Kristi Potter (NREL - Golden, US)*



This proposal aims to close the loop on the traditional flow of knowledge by relating simulation and analysis results to a conceptual model. This new framework will relate relevant pieces of scientific workflow, including analysis results, uncertainty information, and background knowledge, to help to solve the intractable data problem faced by exascale computing by explicitly conveying relationships between complex computational systems, large-scale data, and theoretical scientific concepts.

## 3.30 A critical analysis of evaluation in medical visualisation

*Bernhard Preim (Universität Magdeburg, DE)*



Medical visualisation is typically evaluated anecdotical only. In addition there are some perception-based studies related to shape and depth perception. These are serious quantitative evaluation but they are limited to understanding low level perceptual issues and not the high level cognitive activities like the decision-making and problem solving in diagnosis, treatment planning and medical education. Eye-tracking studies, think about, interaction protocols and long term case studies are needed to better understand what works in medical visualisation.

## 3.31 Mathematical Foundations in my Work

*Gerik Scheuermann (Universität Leipzig, DE)*



In a short position statement for the mathematical foundations panel, I name the areas of mathematics that I have used in the past. From the insight that pretty much all areas of mathematics have been used to some extent for visualization purposes in the literature, I raise the question which areas should be part of a curriculum and which areas are just optional depending on the specific material covered in class. Besides, I also pointed at the problem that the meaning of mathematical concepts behind visualization algorithms does not fit the applications in some cases, leading to unsatisfying results.

## 3.32 Collaborating with Domain Experts

*Marc Streit (Johannes Kepler Universität Linz, AT)*



In the first part of my talk, I summarize what advice researchers and practitioners can get from a theory of visualization. We – as a community – currently provide advice by publishing models and theories, by collecting techniques and methods, and by describing best practices. While this is very useful, it is often not actionable. A less explored possibility is to provide cheat sheets in the form of decision trees that can help practitioners to create effective visualizations. These decision trees could be created as a community effort, underpinned with our models, and carefully annotated. In the second part, I talk about why generalizing design studies is hard, why data and task abstraction is key to create impact in visualization through collaboration with domain experts, and what lessons I've learned in previous collaborations.

## 3.33 Mathematical Foundations of Visualization – Different Kinds

*Holger Theisel (Universität Magdeburg, DE)*



Which mathematical foundations should we expect from Visualization experts? There are two kinds: foundations for all Visualization experts, and foundations that only a few visualization experts work with. This is fine: it is an established way of Data Vis development to constantly discover (not invent) new mathematical theories for Visualization, and to make them useful and applicable for visualize concrete data. Further, visualization experts can and should contribute in developing the foundations We should not wait for the results and "only" visualize them then! This is, however, an approach limited to only a few problems in visualization (perhaps the ones visualization is most matured)

## 3.34 Bringing your research to broad audiences

*Jarke J. van Wijk (TU Eindhoven, NL)*



Having your research published at one of our venues is not an endpoint. It can be highly useful and rewarding to bring it to broader audiences. I describe one of my experiences in this. After having developed the cushion treemap technique (1999), I had a student integrate that in a tool just for hard disk visualization, Sequoiaview. That attracted much attention, and led to generalization of the method and a start-up company, MagnaView. One of their successes was a tool for visualizing high school data, which is used on a large scale. To be succesful in this, dedication to the needs and wants of the audience and careful tuning of presentation and interaction is crucial.

## 3.35 Domain Expert Collaboration: when it went well

*Anna Vilanova (TU Delft, NL)*



I present an example in which the collaboration was a success according to my definition of success.

Key factors:

- New data that the domain experts could not analyze without visualization aid.
- A problem that suits the visualization field and has challenges that are unsolved in the vis community.
- Two vis people: Have a person in the project just focused on the development of general Vis techniques that are inspired but not directly application dependent. Have another person between domains that transforms advances to an adapted framework that can be used by the domain experts.
- Engineering factor needed was limited
- Funding was available, with quite some freedom on how to use it.
- Great respect, and effort to understand each others field.
- Talented, communicative and enthusiastic people involved.

## 3.36 On Visual Abstraction

*Ivan Viola (TU Wien, AT)*



Visual abstraction is a fundamental concept in visual arts and data visualization. While we have an intuitive understanding what the term "visual abstraction" stands for, there is no consensus. Abstract, originating from Latin abstrahere, means drawn away, and is often used in terms like abstract data, abstract class, abstract art, where it represents aspects that are derived from a concrete corresponding object. Abstraction is a process and also an outcome of that process. Visual abstraction is therefore a process of abstraction, where information is transformed into visual representations. We can recognize multiple fundamentally different directions of visual abstractions: geometric abstraction, photometric abstraction, and temporal abstraction.

## 3.37 Vis4Vis: Visualization in Empirical Visualization Research

*Daniel Weiskopf (Universität Stuttgart, DE)*



Appropriate evaluation in visualization research is a longstanding, relevant, and often-discussed issue. My talk focuses on empirical studies with user involvement. I argue that one of the underlying difficulties is the varying role of visualization research: it has facets of engineering and (natural) science, depending on the research objective at hand. We may

adopt study methodology from other fields, such as psychology or HCI, but have to be careful to adapt them to the specific needs of visualization research. One promising direction is the use of data-rich observations that we can acquire during studies in order to obtain more reliable interpretations of empirical studies. For example, we have been witnessing an increased availability and use of physiological sensor information from eye tracking, EEG, and other modalities as well as user logging. Such data-rich empirical studies promise to be especially useful for studies "in the wild" and similar scenarios. However, with the growing availability of large, complex, time-dependent, heterogeneous, and unstructured observational data, we are facing the new challenge of how we can analyze such data. I argue that we need Vis4Vis: visualization as a means of data analysis of empirical study data to advance visualization research.

## 3.38 Data transformations, embeddings, summaries

*Ross Whitaker (University of Utah - Salt Lake City, US)*



Fundamentally, data visualization is the process of placing dabs of ink or color on a 2D plane. However, the complexity of data is increasing so that we see large numbers of instances, dimensions, paramaters, etc. Such data surpases what can readily shown on a 2D or 3D display. One solution to this challenge is the development of better or more complex interfaces, that include, for instance, linked views, large displays, dynamic visualizations, and sophisticated user interactions. The alternative and complementary approach is to develop sets of mathematical and statistical tools to transform, map, or summarize data and reduce its complexity so that visualization and understanding of large, complex becomes more feasible. The role of visualization research, in this case, is to identify common use cases and develop methods and tools that can readily be adapted to particular applications.

## 3.39 Trust in Visualization (and what it has to do with Theory)

*Thomas Wischgoll (Wright State University - Dayton, US)*



There are different issues with trust involved when working with domain experts to visualize their data. There may be limitations with the data that require special precautions, such as sensitivity or security limitations. It may have taken a lot of effort to collect or create the data so that a certain level of trust is required for the domain expert to share the data. At the same time, the domain expert needs to be able to trust in the final visualization results. This presentation discusses these issues with trust and what requirements for a theoretical foundation this results in. Furthermore, additional requirements are discussed for user interfaces and other elements within the visualization.

## 3.40 Exploranation

*Anders Ynnerman (Linköping University, SE)*



This presentation discusses visualization approaches to reach broad audiences. The area is wide and includes aspects of infographics, science communication, interfaces for human in the loop applications, and indeed specific visualization for large groups of domain experts. This presentation introduces the confluence of exploratory and explanatory visualization denoted "Exploranation" as a means to reach large user groups with engaging visualization. Examples are give from the field of science communication at public venues and presents derived design principles for interactive installations in museums as well as requirements and challenges for mediated visual science communication. The presentation is concluded with reflections on the need for visualization in human in the loop applications such as autonomous systems and presents a visions for visual cognitive companions.

## 3.41 Using Empirical Results in Practice

*Caroline Ziemkiewicz (Forrester Research Inc., US)*



There is a growing and welcome tendency in the visualization community to reflect on how and why to perform empirical studies, particularly user evaluations. For this process of reflection to be productive, it is necessary to consider the various audiences of empirical results and what they need. One important such audience includes visualization practitioners and designers. Practitioners use empirical research results to support decisions about what techniques to use for an application and how to tell whether a design is effective. Generalizing and making use of results in this way requires a full understanding of the context in which the study was performed: task abstractions, user models, assumptions, and tested requirements. Many common methods of designing and reporting empirical studies in visualization lack this context, particularly in system evaluation and technique comparisons. New approaches and methods are needed to make this context concrete and produce results that are specific enough to be generalized.

# 4 Working groups

There were five working groups for the five central topics, i.e.

- Theory of overall visualization process
- Foundations of evaluation
- Collaboration with domain experts
- Visualization for broad audiences
- Mathematical foundations of visual data analysis

One key target product of the workshop was an edited volume on Foundations of Data Visualization. The five working groups explored their topics and organization for sections

of that planned book. Seminar participants were surveyed before the seminar about their level of interest in each working group topic and assigned to working groups based on those interests. Each working group developed a plan for the creation of their book section over the months following the seminar. By the end of the seminar, working groups had identified chapters and authors for the section, as well as a schedule for authoring and review. Work on the chapters themselves continues after the conclusion of the seminar.

## Participants

- James Ahrens
Los Alamos National Lab., US
- Johanna Beyer
Harvard University – Cambridge, US
- Michael Böttinger
DKRZ Hamburg, DE
- Stefan Bruckner
University of Bergen, NO
- Roxana Bujack
Los Alamos National Laboratory, US
- Hamish Carr
University of Leeds, GB
- Min Chen
University of Oxford, GB
- Leila De Floriani
University of Maryland – College Park, US
- Christoph Garth
TU Kaiserslautern, DE
- Eduard Gröller
TU Wien, AT
- Hans Hagen
TU Kaiserslautern, DE
- Charles D. Hansen
University of Utah – Salt Lake City, US
- Helwig Hauser
University of Bergen, NO
- Hans-Christian Hege
Zuse Institute Berlin, DE
- Nathalie Henry Riche
Microsoft Research – Redmond, US
- Mario Hlawitschka
Hochschule Leipzig, DE
- Ingrid Hotz
Linköping University, SE
- Christopher R. Johnson
University of Utah – Salt Lake City, US
- Alark Joshi
University of San Francisco, US
- Gordon Kindlmann
University of Chicago, US
- Helen-Nicole Kostis
USRA/GESTAR SVS NASA/GSFC, US
- Barbora Kozlíková
Masaryk University – Brno, CZ
- David H. Laidlaw
Brown University – Providence, US
- Heike Leitte
TU Kaiserslautern, DE
- Ross Maciejewski
Arizona State University – Tempe, US
- Georgeta Elisabeta Marai
University of Illinois – Chicago, US
- Kresimir Matkovic
VRVis – Wien, AT
- Laura A. McNamara
Sandia National Labs – Albuquerque, US
- Silvia Miksch
TU Wien, AT
- Torsten Möller
Universität Wien, AT
- Daniela Oelke
Siemens AG – München, DE
- Kristi Potter
NREL – Golden, US
- Bernhard Preim
Universität Magdeburg, DE
- Penny Rheingans
University of Maryland, Baltimore County, US
- Gerik Scheuermann
Universität Leipzig, DE
- Marc Streit
Johannes Kepler Universität Linz, AT
- Holger Theisel
Universität Magdeburg, DE
- Jarke J. van Wijk
TU Eindhoven, NL
- Amitabh Varshney
University of Maryland – College Park, US
- Maria Velez-Rojas
CA Technologies – Santa Clara, US
- Anna Vilanova
TU Delft, NL
- Ivan Viola
TU Wien, AT
- Daniel Weiskopf
Universität Stuttgart, DE
- Ross Whitaker
University of Utah – Salt Lake City, US
- Thomas Wischgoll
Wright State University – Dayton, US
- Anders Ynnerman
Linköping University, SE
- Caroline Ziemkiewicz
Forrester Research Inc., US

Report from Dagstuhl Seminar 18051

# Proof Complexity

**Edited by**

**Albert Atserias[1], Jakob Nordström[2], Pavel Pudlák[3], and Rahul Santhanam[4]**

1 **UPC – Barcelona, ES,** atserias@cs.upc.edu
2 **KTH Royal Institute of Technology – Stockholm, SE,** jakobn@kth.se
3 **The Czech Academy of Sciences – Prague, CZ,** pudlak@math.cas.cz
4 **University of Oxford, GB,** rahul.santhanam@cs.ox.ac.uk

## Abstract

The study of proof complexity was initiated in [Cook and Reckhow 1979] as a way to attack the P vs. NP problem, and in the ensuing decades many powerful techniques have been discovered for analyzing different proof systems. Proof complexity also gives a way of studying subsystems of Peano Arithmetic where the power of mathematical reasoning is restricted, and to quantify how complex different mathematical theorems are measured in terms of the strength of the methods of reasoning required to establish their validity. Moreover, it allows to analyse the power and limitations of satisfiability algorithms (SAT solvers) used in industrial applications with formulas containing up to millions of variables.

During the last 10–15 years the area of proof complexity has seen a revival with many exciting results, and new connections have also been revealed with other areas such as, e.g., cryptography, algebraic complexity theory, communication complexity, and combinatorial optimization. While many longstanding open problems from the 1980s and 1990s still remain unsolved, recent progress gives hope that the area may be ripe for decisive breakthroughs. This workshop, gathering researchers from different strands of the proof complexity community, gave opportunities to take stock of where we stand and discuss the way ahead.

**Seminar** January 28–February 2, 2018 – https://www.dagstuhl.de/18051
**2012 ACM Subject Classification** Theory of computation → Proof complexity
**Keywords and phrases** bounded arithmetic, computational complexity, logic, proof complexity, satisfiability algorithms
**Digital Object Identifier** 10.4230/DagRep.8.1.124
**Edited in cooperation with** Marc Vinyals

## 1 Executive Summary

*Albert Atserias*
*Jakob Nordström*
*Pavel Pudlák*
*Rahul Santhanam*



This workshop brought together the whole proof complexity community spanning from Frege proof systems and circuit-inspired lower bounds via geometric and algebraic proof systems all the way to bounded arithmetic. In this executive summary, we first give an overview of proof complexity, and then describe the goals of the seminar week. Finally, we discuss the relation to previous workshops and conferences.

## Topic of the Seminar

Ever since the groundbreaking NP-completeness paper of Cook [18], the problem of deciding whether a given propositional logic formula is satisfiable or not has been on centre stage in theoretical computer science. During the last two decades, SATISFIABILITY has also developed from a problem of mainly theoretical interest into a practical approach for solving applied problems. Although all known Boolean satisfiability solvers (SAT solvers) have exponential running time in the worst case, enormous progress in performance has led to satisfiability algorithms becoming a standard tool for solving large-scale problems in, for example, hardware and software verification, artificial intelligence, bioinformatics, operations research, and sometimes even pure mathematics.

The study of proof complexity originated with the seminal paper of Cook and Reckhow [19]. In its most general form, a proof system for a formal language $L$ is a predicate $P(x, \pi)$, computable in time polynomial in the sizes $|x|$ and $|\pi|$ of the input, and having the property that for all $x \in L$ there exists a string $\pi$ (a *proof*) for which $P(x, \pi)$ evaluates to true, whereas for any$x \notin L$ it should hold for all strings $\pi$ that $P(x, \pi)$ evaluates to false. A proof system is said to be polynomially bounded if for every $x \in L$ there exists a proof $\pi_x$ for $x$ that has size at most polynomial in $|x|$. A *propositional proof system* is a proof system for the language of tautologies in propositional logic, i.e., for formulas that always evaluate to true no matter how the values true and false are assigned to variables in the formula.

From a theoretical point of view, one important motivation for proof complexity is the intimate connection with the fundamental problem of P versus NP. Since NP is exactly the set of languages with polynomially bounded proof systems, and since TAUTOLOGY can be seen to be the dual problem of SATISFIABILITY, we have the famous theorem of [19] that NP = coNP if and only if there exists a polynomially bounded propositional proof system. Thus, if it could be shown that there are no polynomially bounded proof systems for tautologies, P $\neq$ NP would follow as a corollary since P is closed under complement. One way of approaching this problem is to study stronger and stronger proof systems and try to prove superpolynomial lower bounds on proof size. However, although great progress has been made in the last couple of decades for a variety of proof systems, this goal still appears very distant.

A second theoretical motivation is that simple propositional proof systems provide analogues of subsystems of Peano Arithmetic where the power of mathematical reasoning is restricted. Of particular interest here are various bounded arithmetic systems, which in some sense are intended to capture feasible/polynomial-time reasoning. Proving strong lower bounds on propositional logic encodings of some combinatorial principle, say, in a propositional proof system can in this way show that establishing the validity of this principle requires more powerful mathematics than what is provided by the corresponding subsystem of Peano Arithmetic. One can thus quantify how "deep" different mathematical truths are, as well as shed light on the limits of our (human, rather than automated) proof techniques. At the same time, since it is an empirically verified fact that low-complexity proofs generalize better and are often more constructive, classifying which truths have feasible proofs is also a way to approach the classification of algorithmic problems by their computational complexity. The precise sense in which this can be formalized into a tool for the complexity theorist is one of the goals of bounded arithmetic.

A third prominent motivation for the study of proof complexity is also algorithmic but of a more practical nature. As was mentioned above, designing efficient algorithms for proving tautologies–or, equivalently, testing satisfiability–is a very important problem not only in the theory of computation but also in applied research and industry. All SAT solvers, regardless

of whether they produce a written proof or not, explicitly or implicitly define a system in which proofs are searched for and rules which determine what proofs in this system look like. Proof complexity analyses what it takes to simply write down and verify the proofs that such a solver might find, ignoring the computational effort needed to actually find them. Thus, a lower bound for a proof system tells us that any algorithm, even an optimal (non-deterministic) one magically making all the right choices, must necessarily use at least the amount of a certain resource specified by this bound. In the other direction, theoretical upper bounds on some proof complexity measure give us hope of finding good proof search algorithms with respect to this measure, provided that we can design algorithms that search for proofs in the system in an efficient manner.

The field of proof complexity also has rich connections to algorithmic analysis, combinatorial optimization, cryptography, artificial intelligence, and mathematical logic. A few good sourcesproviding more details are [6, 17, 47].

## A Very Selective Survey of Proof Complexity

Any propositional logic formula can be converted to a formula in conjunctive normal form (CNF) that is only linearly larger and is unsatisfiable if and only if the original formula is a tautology. Therefore, any sound and complete system that certifies the unsatisfiability of CNF formulas can be considered as a general propositional proof system.

The extensively studied *resolution* proof system, which appeared in [9] and began to be investigated in connection with automated theorem proving in the 1960s [21, 22, 48], is such a system where one derives new disjunctive clauses from an unsatisfiable CNF formula until an explicit contradiction is reached. Despite the apparent simplicity of resolution, the first superpolynomial lower bounds on proof size were obtained only after decades of study in 1985 [33], after which truly exponential size lower bounds soon followed in [15, 52]. It was shown in [8] that these lower bounds can be established by instead studying the *width* of proofs, i.e., the maximal size of clauses in the proofs, and arguing that any resolution proof for a certain formula must contain a large clause. It then follows by a generic argument that any such proof must also consist of very many clauses. Later research has led to a well-developed machinery for showing width lower bounds, and hence also size lower bounds, for resolution.

The more general proof system *polynomial calculus (PC)*, introduced in [1, 16],[1] instead uses algebraic geometry to reason about SAT. In polynomial calculus clauses are translated to multilinear polynomials over some (fixed) field, and a CNF formula $F$ is shown to be unsatisfiable by proving that there is no common root for the polynomials corresponding to all the clauses, or equivalently that the multiplicative identity 1 lies in the ideal generated by these polynomials. Here the size of a proof is measured as the number of monomials in a proof when all polynomials are expanded out as linear combinations of monomials, and the width of a clause corresponds to the (total) *degree* of the polynomial representing the clause. It can be shown that PC is at least as strong as resolution with respect to both size and width/degree, and there are families of formulas for which PC is exponentially stronger.

In the work [36], which served, interestingly enough, as a precursor to [8], it was shown that strong lower bounds on the degree of polynomial calculus proofs are sufficient to establish strong size lower bounds. In contrast to the situation for resolution after [8], however, this

[1] Expert readers will note that we do not distinguish between PC [16] and PCR [1] below due to space constraints.

has not been followed by a corresponding development of a generally applicable machinery for proving degree lower bounds. For fields of characteristic distinct from 2 it is sometimes possible to obtain lower bounds by doing an affine transformation from $\{0,1\}$ to the "Fourier basis" $\{-1,+1\}$, an idea that seems to have appeared first in [13, 28]. For fields of arbitrary characteristic a powerful technique for general systems of polynomial equations was developed in [2], which when restricted to CNF formulas $F$ yields that polynomial calculus proofs require high degree if the corresponding clause-variable incidence graphs $G(F)$ are good enough bipartite expander graphs. There are several provably hard formula families for which this criterion fails to apply, however, and even more formulas that are believed to be hard for both resolution and PC, but where lower bounds are only known for the former proof system and not the latter.

Another proof system that has been the focus of much research is *cutting planes (CP)*, which was introduced in [20] as a way of formalizing the integer linear programming algorithm in [14, 27]. Here the disjunctive clauses in a CNF formula are translated to linear inequalities, and these linear inequalities are then manipulated to derive a contradiction. Thus, questions about the satifiability of Boolean formulas are reduced to the geometry of polytopes over the real numbers. Cutting planes is easily seen to be as least as strong as resolution, since a CP proof can mimic any resolution proof line by line. An intriguing fact is that encodings of the *pigeonhole principle*, which are known to be hard to prove for resolution [33] and many other proof systems, are very easy to prove in cutting planes. It follows from this that not only is cutting planes never worse than resolution, but it can be exponentially stronger.

Exponential lower bounds on proof length for cutting planes were first proven in [10] for the restricted subsystem CP$^*$, where all coefficients in the linear inequalities can be at most polynomial in the formula size, and were later extended to general CP in [34, 44]. The proof technique in [44] is very specific, however, in that it works by *interpolating* monotone Boolean circuits for certain problems from CP proofs of related formulas with a very particular structure, and then appealing to lower bounds in circuit complexity. A longstanding open problem is to develop techniques that would apply to other formula families. For example, establishing that randomly sampled $k$-CNF formulas are hard to refute for CP, or that CP cannot efficiently prove the fact that the sum of all vertex degrees in an undirected graph is even (encoded in so-called *Tseitin formulas*), would constitute major breakthroughs.

We remark that there are also other proof systems inspired by linear and semidefinite programming, e.g., in [38, 39, 50], which are somewhat similar to but incomparable with cutting planes, and a deeper understanding of which appear even more challenging. Some notable early papers in proof complexity investigating these so-called *semialgebraic proof systems* were published around the turn of the millennium in [30, 31, 45], but then this area of research seems to have gone dormant. In the last few years, these proof systems have made an exciting reemergence in the context of hardness of approximation, revealing unexpected and intriguing connections between approximation and proof complexity. A precursor to this is the work by Schoenebeck [49], which gave strong integrality gaps in the so-called Lasserre SDP hierarchy using results from proof complexity. These results were later realized to be a rediscovery of results by Grigoriev [29] proving degree lower bounds for what he called the *Positivstellensatz Calculus* [31]. More recently we have the work of Barak et al. [4], which was the first to explicitly point out this intriguing connection between approximability and proof complexity. Following this paper, several papers have appeared that continue the fruitful exploration of the interplay between approximability and proof complexity. Results from this area also appeared in the invited talk of Boaz Barak at the International Congress of Mathematicians in 2014 (see [5]).

The paper [19] initiated research in proof complexity focused on a more general and powerful family of propositional proof systems called *Frege systems*. Such systems consist of a finite implicationally complete set of axioms and inference rules (let us say over connectives AND, OR, and NOT for concreteness), where new formulas are derived by substitution into the axioms and inference rules. Various forms of Frege systems (also called *Hilbert systems*) typically appear in logic textbooks, and typically the exact definitions vary. Such distinctions do not matter for our purposes, however—it was shown in [19] that all such systems are equivalent up to an at most polynomial blow-up in the proof size.

Frege systems are well beyond what we can prove nontrivial lower bounds for; the situation is similar to the problem of proving lower bound on the size of Boolean circuits. Therefore restricted versions of Frege systems have been studied. One natural restriction is to allow unbounded fan-in AND-OR formulas (where negations appear only in front of atomic variables) but to require that all formulas appearing in a proof have bounded depth (i.e., a bounded number of alternations between AND and OR). Such a model is an analogue of the bounded-depth circuits studied in circuit complexity, but first arose in the context of bounded first-order arithmetic in logic [12, 41]. For such *bounded-depth Frege systems* exponential lower bounds on proof size were obtained in [37, 42], but these lower bounds only work for depth smaller than $\log \log n$. This depth lower bound was very recently improved to $\sqrt{\log n}$ in [43], but in terms of the size lower bound this recent result is much weaker. By comparison, for the corresponding class in circuit complexity strong size lower bounds are known all the way up to depth $\log n / \log \log n$. Also, if one extends the set of connectives with exclusive or (also called parity) to obtain *bounded-depth Frege with parity gates*, then again no lower bounds are known, although strong lower bounds have been shown for the analogous class in circuit complexity [46, 51].

The quest for lower bounds for bounded-depth Frege systems and beyond are mainly motivated by the P vs. NP problem. Regarding connections to SAT solving, it is mostly weaker proof systems such as resolution, polynomial calculus, and cutting planes that are of interest, whereas the variants of Frege systems discussed above do not seem to be suitable foundations for SAT solvers. The issue here is that not only do we want our proof system to be as powerful as possible, i.e., having short proofs for the formulas under consideration, but we also want to be able to *find these proofs efficiently.*

We quantify this theoretically by saying that a proof system is *automatizable* if there is an algorithm that finds proofs in this system in time polynomial in the length of an optimal proof. This seems to be the right notion: If there is no short proof of a formula in the system, then we cannot expect any algorithm to find a proof quickly, but if there is a short proof to be found we want an algorithm that is competitive with respect to the length of such a proof. Unfortunately, there seems to be a trade-off here in the sense that if a proof system is sufficiently powerful, then it is not automatizable. For instance, bounded-depth Frege systems are not automatizable under plausible computational complexity assumptions [11]. However, analogous results have later been shown also for resolution [3], and yet proof search is implemented successfully in this proof system in practice. This raises intriguing questions that seem to merit further study.

## Goals of the Seminar

There is a rich selection of open problems that could be discussed at a workshop focused on proof complexity. Below we just give a few samples of such problems that came up during

the workshop–it should be emphasized that this list is very far from exhaustive and is only intended to serve as an illustration.

For starters, there are a number of NP-complete problems for which we would like to understand the hardness with respect to polynomial calculus and other algebraic proof systems. For the problem of cliques of constant size $k$ in graphs, there is an obvious polynomial-time algorithm (since only $\binom{n}{k} \leq n^k$ possible candidate cliques need to be checked). Whether this brute-force algorithm is optimal or not is a deep question with connections to fixed-parameter tractability and parameterized proof complexity. This is completely open for polynomial calculus, and even for resolution. The ultimate goal here would be to prove average-case lower bounds for $k$-clique formulas over Erdős–Rényi random graphs $G(n, p)$ with edge probability just below the threshold $p = n^{-2/(k-1)}$ for the appearance of $k$-cliques.

In contrast to the clique problem, graph colouring is NP-complete already for a constant number 3 of colours. If we believe that $\mathsf{P} \neq \mathsf{NP}$, then, in particular, it seems reasonable to expect that this problem should be hard for polynomial calculus. No such results have been known, however. On the contrary, in the papers [23, 24, 25] recognized with the *INFORMS Computing Society Prize 2010*, the authors report that they used algebraic methods formalizable in polynomial calculus that "successfully solved graph problem instances having thousands of nodes and tens of thousands of edges" and that they could not find hard instances for these algorithms. This is very surprising. For resolution, it was shown in [7] that random graphs with the right edge density are exponentially hard to deal with, and it seems likely that the same should hold also for polynomial calculus. This appears to be a very challenging problem, however, but we hope that techniques from [2, 40] can be brought to bear on it.

For cutting planes, a longstanding open problem is to prove lower bounds for random $k$-CNF formulas or Tseitin formulas over expander graphs. An interesting direction in the last few years has been the development of new techniques for size-space trade-offs, showing that if short cutting planes proofs do exist, such proofs must at least have high space complexity in that they require a lot of memory to be verified. Such results were first obtained via a somewhat unexpected connection to communication complexity in [35], and have more recently been strengthened in [26, 32].

Admittedly, proving lower bounds for bounded-depth Frege systems and beyond is another formidable challenge, and it only seems prudent to say that this is a high-risk proposal. However, the very recent, and exciting, progress in [43] give hope that new techniques might be developed to attack also this problem.

## Relation to Previous Dagstuhl Seminars

The area of proof complexity has a large intersection with computational complexity theory, and are two recurring workshops at Dagstuhl dedicated to complexity theory broadly construed, namely *Computational Complexity of Discrete Problems* and *Algebraic Methods in Computational Complexity*. However, these two workshops have had very limited coverage of topics related to proof complexity in the past.

On the more applied side, there have been two workshops *SAT and Interactions* and *Theory and Practice of SAT Solving* that have explored the connections between computational complexity and more applied satisfiability algorithms as used in industry (so-called SAT solvers). These workshops have focused on very weak proof systems, however, which are the ones that are of interest in connection to SAT solving, but have not made any connections to stronger proof systems or to bounded arithmetic.

Although proof complexity has turned out to have deep connections to both complexity theory and SAT solving, proof complexity is an interesting and vibrant enough area to merit a seminar week in its own right. This workshop at Dagstuhl provided a unique opportunity for the community to meet during a full week focusing on the latest news in various subareas and major challenges going forward.

## References


1. Michael Alekhnovich, Eli Ben-Sasson, Alexander A. Razborov, and Avi Wigderson. Space complexity in propositional calculus. *SIAM Journal on Computing*, 31(4):1184–1211, 2002. Preliminary version in *STOC '00*.
2. Michael Alekhnovich and Alexander A. Razborov. Lower bounds for polynomial calculus: Non-binomial case. *Proceedings of the Steklov Institute of Mathematics*, 242:18–35, 2003. Available at http://people.cs.uchicago.edu/~razborov/files/misha.pdf. Preliminary version in *FOCS '01*.
3. Michael Alekhnovich and Alexander A. Razborov. Resolution is not automatizable unless W[P] is tractable. *SIAM Journal on Computing*, 38(4):1347–1363, October 2008. Preliminary version in *FOCS '01*.
4. Boaz Barak, Fernando G. S. L. Brandão, Aram Wettroth Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC '12)*, pages 307–326, May 2012.
5. Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. In *Proceedings of International Congress of Mathematicians (ICM)*, 2014.
6. Paul Beame. Proof complexity. In Steven Rudich and Avi Wigderson, editors, *Computational Complexity Theory*, volume 10 of *IAS/Park City Mathematics Series*, pages 199–246. American Mathematical Society, 2004.
7. Paul Beame, Joseph C. Culberson, David G. Mitchell, and Cristopher Moore. The resolution complexity of random graph $k$-colorability. *Discrete Applied Mathematics*, 153(1-3):25–47, December 2005.
8. Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow—resolution made simple. *Journal of the ACM*, 48(2):149–169, March 2001. Preliminary version in *STOC '99*.
9. Archie Blake. *Canonical Expressions in Boolean Algebra*. PhD thesis, University of Chicago, 1937.
10. María Bonet, Toniann Pitassi, and Ran Raz. Lower bounds for cutting planes proofs with small coefficients. In *Proceedings of the 27th Annual ACM Symposium on Theory of Computing (STOC '95)*, pages 575–584, May 1995.
11. María Luisa Bonet, Carlos Domingo, Ricard Gavaldà, Alexis Maciel, and Toniann Pitassi. Non-automatizability of bounded-depth Frege proofs. In *Proceedings of the 14th Annual IEEE Conference on Computational Complexity (CCC '99)*, pages 15–23, May 1999.
12. Samuel R. Buss. *Bounded Arithmetic*. Bibliopolis, Naples, 1986. Revision of PhD thesis.
13. Samuel R. Buss, Dima Grigoriev, Russell Impagliazzo, and Toniann Pitassi. Linear gaps between degrees for the polynomial calculus modulo distinct primes. *Journal of Computer and System Sciences*, 62(2):267–289, March 2001. Preliminary version in *CCC '99*.
14. Vašek Chvátal. Edmonds polytopes and a hierarchy of combinatorial problems. *Discrete Mathematics*, 4(1):305–337, 1973.
15. Vašek Chvátal and Endre Szemerédi. Many hard examples for resolution. *Journal of the ACM*, 35(4):759–768, October 1988.
16. Matthew Clegg, Jeffery Edmonds, and Russell Impagliazzo. Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC '96)*, pages 174–183, May 1996.

**17** Peter Clote and Evangelos Kranakis. *Boolean Functions and Computation Models*. Springer, 2002.

**18** Stephen A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing (STOC '71)*, pages 151–158, 1971.

**19** Stephen A. Cook and Robert Reckhow. The relative efficiency of propositional proof systems. *Journal of Symbolic Logic*, 44(1):36–50, March 1979.

**20** William Cook, Collette Rene Coullard, and György Turán. On the complexity of cutting-plane proofs. *Discrete Applied Mathematics*, 18(1):25–38, November 1987.

**21** Martin Davis, George Logemann, and Donald Loveland. A machine program for theorem proving. *Communications of the ACM*, 5(7):394–397, July 1962.

**22** Martin Davis and Hilary Putnam. A computing procedure for quantification theory. *Journal of the ACM*, 7(3):201–215, 1960.

**23** Jesús A. De Loera, Jon Lee, Peter N. Malkin, and Susan Margulies. Hilbert's Nullstellensatz and an algorithm for proving combinatorial infeasibility. In *Proceedings of the 21st International Symposium on Symbolic and Algebraic Computation (ISSAC '08)*, pages 197–206, July 2008.

**24** Jesús A. De Loera, Jon Lee, Peter N. Malkin, and Susan Margulies. Computing infeasibility certificates for combinatorial problems through Hilbert's Nullstellensatz. *Journal of Symbolic Computation*, 46(11):1260–1283, November 2011.

**25** Jesús A. De Loera, Jon Lee, Susan Margulies, and Shmuel Onn. Expressing combinatorial problems by systems of polynomial equations and Hilbert's Nullstellensatz. *Combinatorics, Probability and Computing*, 18(04):551–582, July 2009.

**26** Susanna F. de Rezende, Jakob Nordström, and Marc Vinyals. How limited interaction hinders real communication (and what it means for proof and circuit complexity). In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS '16)*, pages 295–304, October 2016.

**27** Ralph E. Gomory. An algorithm for integer solutions of linear programs. In R.L. Graves and P. Wolfe, editors, *Recent Advances in Mathematical Programming*, pages 269–302. McGraw-Hill, New York, 1963.

**28** Dima Grigoriev. Tseitin's tautologies and lower bounds for Nullstellensatz proofs. In *Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science (FOCS '98)*, pages 648–652, November 1998.

**29** Dima Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1–2):613–622, May 2001.

**30** Dima Grigoriev, Edward A. Hirsch, and Dmitrii V. Pasechnik. Complexity of semialgebraic proofs. *Moscow Mathematical Journal*, 2(4):647–679, 2002. Preliminary version in *STACS '02*.

**31** Dima Grigoriev and Nicolai Vorobjov. Complexity of Null- and Positivstellensatz proofs. *Annals of Pure and Applied Logic*, 113(1–3):153–160, December 2001.

**32** Mika Göös and Toniann Pitassi. Communication lower bounds via critical block sensitivity. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC '14)*, pages 847–856, May 2014.

**33** Armin Haken. The intractability of resolution. *Theoretical Computer Science*, 39(2-3):297–308, August 1985.

**34** Armin Haken and Stephen A. Cook. An exponential lower bound for the size of monotone real circuits. *Journal of Computer and System Sciences*, 58(2):326–335, 1999.

**35** Trinh Huynh and Jakob Nordström. On the virtue of succinct proofs: Amplifying communication complexity hardness to time-space trade-offs in proof complexity (Extended abstract). In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC '12)*, pages 233–248, May 2012.

**36** Russell Impagliazzo, Pavel Pudlák, and Jiří Sgall. Lower bounds for the polynomial calculus and the Gröbner basis algorithm. *Computational Complexity*, 8(2):127–144, 1999.

**37** Jan Krajíček, Pavel Pudlák, and Alan R. Woods. An exponential lower bound to the size of bounded depth Frege proofs of the pigeonhole principle. *Random Structures and Algorithms*, 7(1):15–40, 1995. Preliminary version in *STOC '92*.

**38** Jean B. Lasserre. An explicit exact SDP relaxation for nonlinear 0-1 programs. In *Proceedings of the 8th International Conference on Integer Programming and Combinatorial Optimization (IPCO '01)*, volume 2081 of *Lecture Notes in Computer Science*, pages 293–303. Springer, June 2001.

**39** László Lovász and Alexander Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991.

**40** Mladen Mikša and Jakob Nordström. A generalized method for proving polynomial calculus degree lower bounds. In *Proceedings of the 30th Annual Computational Complexity Conference (CCC '15)*, volume 33 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 467–487, June 2015.

**41** Jeff B. Paris and Alex J. Wilkie. Counting problems in bounded arithmetic. In *Methods in Mathematical Logic: Proceedings of the 6th Latin American Symposium on Mathematical Logic*, volume 1130 of *Lecture Notes in Mathematics*, pages 317–340. Springer, 1985.

**42** Toniann Pitassi, Paul Beame, and Russell Impagliazzo. Exponential lower bounds for the pigeonhole principle. *Computational Complexity*, 3:97–140, 1993. Preliminary version in *STOC '92*.

**43** Toniann Pitassi, Benjamin Rossman, Rocco Servedio, and Li-Yang Tan. Poly-logarithmic Frege depth lower bounds. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC '16)*, pages 644–657, June 2016.

**44** Pavel Pudlák. Lower bounds for resolution and cutting plane proofs and monotone computations. *Journal of Symbolic Logic*, 62(3):981–998, September 1997.

**45** Pavel Pudlák. On the complexity of propositional calculus. In S. Barry Cooper and John K. Truss, editors, *Sets and Proofs*, volume 258 of *London Mathematical Society Lecture Note Series*, pages 197–218. Cambridge University Press, 1999.

**46** Alexander A. Razborov. Lower bounds on the size of bounded depth networks over a complete basis with logical addition. *Matematicheskie Zametki*, 41(4):598–607, 1987. English Translation in *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987.

**47** Alexander A. Razborov. Proof complexity and beyond. *ACM SIGACT News*, 47(2):66–86, June 2016.

**48** John Alan Robinson. A machine-oriented logic based on the resolution principle. *Journal of the ACM*, 12(1):23–41, January 1965.

**49** Grant Schoenebeck. Linear level Lasserre lower bounds for certain $k$-CSPs. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS '08)*, pages 593–602, October 2008.

**50** Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3:411–430, 1990.

**51** Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC '87)*, pages 77–82, 1987.

**52** Alasdair Urquhart. Hard examples for resolution. *Journal of the ACM*, 34(1):209–219, January 1987.

## 2 Contents

# 3 Overview of Presentations Given During the Seminar Week

In this section we list the talks given during the seminar week. As can be seen from a comparison with Section 1, a number of presentations could report progress on long-standing open problems.

In addition to the list of "official" presentations below, there were also a number of more informal presentations and discussions on various topics (including, but not limited to, the open problems mentioned in Section 4).

## 3.1 Some Classic SOS Gems with Proofs

*Albert Atserias (UPC – Barcelona, ES)*



This will be a blackboard lecture-like talk in which I will define the version of Sums-of-Squares (SOS) proof that I want to discuss, and cover the proofs of two beautiful results about it in (an usual amount of?) detail. The first gem is a surprising new result of Berkholz [1], with an equally surprising simple proof, that shows that SOS simulates Polynomial Calculus over the reals with Boolean-valued variables. The second gem is the beautiful construction of Grigoriev [2], as rediscovered by Schoenebeck [3], for showing that systems of parity equations that are hard for resolution are also hard for SOS.

### References

1 Christoph Berkholz: *The Relation between Polynomial Calculus, Sherali-Adams, and Sum-of-Squares Proofs.* STACS 2018: 11:1–11:14
2 Dima Grigoriev: *Tseitin's Tautologies and Lower Bounds for Nullstellensatz Proofs.* FOCS 1998: 648–652
3 Grant Schoenebeck: *Linear Level Lasserre Lower Bounds for Certain k-CSPs.* FOCS 2008: 593–602

## 3.2 Hard Principles from Bounded Arithmetic

*Arnold Beckmann (Swansea University, GB)*



This talk is intended as a second tutorial on Bounded Arithmetic following that of Neil Thapen. It will focus on how Bounded Arithmetic is useful for obtaining hard principles for propositional proof systems. We will touch on reflection principles and related techniques, and demonstrate their usefulness with a few examples. The main part of the tutorial will concentrate on total NP search problems and their relation to Bounded Arithmetic. We will review recent characterisations of classes of total NP search problems whose totality can be proven in certain Bounded Arithmetic theories, and demonstrate through examples how complete problems for such classes lead to hard problems for propositional proof systems corresponding to Bounded Arithmetic theories.

## 3.3 What's Different in QBF from Propositional Proof Complexity?

*Olaf Beyersdorff (University of Leeds, GB)*


**Main reference** Olaf Beyersdorff, Joshua Blinkhorn: "Genuine Lower Bounds for QBF Expansion", in Proc. of the 35th Symposium on Theoretical Aspects of Computer Science, STACS 2018, February 28 to March 3, 2018, Caen, France, LIPIcs, Vol. 96, pp. 12:1–12:15, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2018.
**URL** http://dx.doi.org/10.4230/LIPIcs.STACS.2018.12

The aim of the talk is to discuss QBF proof complexity in comparison to propositional proof complexity. In particular, I will talk about different ideas for QBF resolution systems, the hard formulas we currently have, what is a genuine QBF lower bound and what techniques we have to show them. As an example of a genuine lower bound I will explain the size-cost-capacity technique [1].

### References

1 Olaf Beyersdorff, Joshua Blinkhorn, Luke Hinde: *Size, Cost and Capacity: A Semantic Technique for Hard Random QBFs.* ITCS 2018: 9:1–9:18

## 3.4 Clique Is Hard on Average for Regular Resolution

*Ilario Bonacina (UPC – Barcelona, ES)*


**Joint work of** Albert Atserias, Ilario Bonacina, Susanna de Rezende, Massimo Lauria, Jakob Nordström, Alexander Razborov
**Main reference** Albert Atserias, Ilario Bonacina, Susanna F. de Rezende, Massimo Lauria, Jakob Nordström, Alexander A. Razborov: "Clique is hard on average for regular resolution", in Proc. of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018, pp. 866–877, ACM, 2018.
**URL** http://dx.doi.org/10.1145/3188745.3188856

Deciding whether a graph $G$ with $n$ vertices has a $k$-clique is one of the most basic computational problems on graphs. In this work we show that certifying $k$-clique-freeness of Erdős–Rényi random graphs is hard for regular resolution. More precisely we show that for $k \ll \sqrt{n}$ regular resolution asymptotically almost surely requires length $n^{\Omega(k)}$ to establish that an Erdős–Rényi random graph (with appropriate edge density) does not contain a $k$-clique. This asymptotically optimal result implies unconditional lower bounds on the running time of several state-of-the-art algorithms used in practice.

## 3.5 Proof Complexity Lower Bounds from Algebraic Circuit Complexity

*Michael A. Forbes (University of Illinois – Urbana-Champaign, US)*


**Joint work of** Michael A. Forbes, Amir Shpilka, Iddo Tzameret, Avi Wigderson
**Main reference** Michael A. Forbes, Amir Shpilka, Iddo Tzameret, Avi Wigderson: "Proof Complexity Lower Bounds from Algebraic Circuit Complexity", in Proc. of the 31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan, LIPIcs, Vol. 50, pp. 32:1–32:17, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2016.
**URL** http://dx.doi.org/10.4230/LIPIcs.CCC.2016.32

We give upper and lower bounds on the power of subsystems of the Ideal Proof System (IPS), the algebraic proof system recently proposed by Grochow and Pitassi [1], where the circuits comprising the proof come from various restricted algebraic circuit classes. This mimics an established research direction in the boolean setting for subsystems of Extended Frege proofs, where proof-lines are circuits from restricted boolean circuit classes. Except one, all of the subsystems considered in this paper can simulate the well-studied Nullstellensatz proof system, and prior to this work there were no known lower bounds when measuring proof size by the algebraic complexity of the polynomials (except with respect to degree, or to sparsity).

We give two general methods of converting certain algebraic lower bounds into proof complexity ones. Our methods require stronger notions of lower bounds, which lower bound a polynomial as well as an entire family of polynomials it defines. Our techniques are reminiscent of existing methods for converting boolean circuit lower bounds into related proof complexity results, such as feasible interpolation. We obtain the relevant types of lower bounds for a variety of classes (sparse polynomials, depth-3 powering formulas, read-once oblivious algebraic branching programs, and multilinear formulas), and infer the relevant proof complexity results. We complement our lower bounds by giving short refutations of the previously-studied subset-sum axiom using IPS subsystems, allowing us to conclude strict separations between some of these subsystems.

### References

1 Joshua A. Grochow, Toniann Pitassi: Circuit Complexity, Proof Complexity, and Polynomial Identity Testing. FOCS 2014: 110–119

## 3.6 On Small-Depth Frege Proofs for Tseitin for Grids

*Johan Hastad (KTH Royal Institute of Technology – Stockholm, SE)*


**Main reference** Johan Håstad: "On Small-Depth Frege Proofs for Tseitin for Grids", in Proc. of the 58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017, pp. 97–108, IEEE Computer Society, 2017.
**URL** http://dx.doi.org/10.1109/FOCS.2017.18

We prove a lower bound on the size of a small depth Frege refutation of the Tseitin contradiction on the grid. We conclude that polynomial size such refutations must use formulas of almost logarithmic depth.

## 3.7 Introduction to Semialgebraic Proof Systems

*Edward A. Hirsch (Steklov Institute – St. Petersburg, RU)*



In this tutorial, I will define semialgebraic proof systems, explain how they work, and survey main results in the area.

## 3.8 Random Formulas and Interpolation in Cutting Planes

*Pavel Hrubes (The Czech Academy of Sciences – Prague, CZ)*


**Joint work of** Pavel Hrubes, Pavel Pudlák
**Main reference** Pavel Hrubes, Pavel Pudlák: "Random Formulas, Monotone Circuits, and Interpolation", in Proc. of the 58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017, pp. 121–131, IEEE Computer Society, 2017.
**URL** https://doi.org/10.1109/FOCS.2017.20

I will discuss the interpolation technique and how it can be adapted to prove new lower bounds for the Cutting Planes proof system. This includes the weak Bit Pigeon Hole Principle and random $\log n$-CNFs.

## 3.9 Parameter-free Bounded Induction

*Emil Jerabek (The Czech Academy of Sciences – Prague, CZ)*



We will have a look at some fragments of bounded arithmetic axiomatized by induction and polynomial induction schemata without parameters.

## 3.10 Bounded-depth Frege with Parity Gates and Subsystems Thereof

*Leszek Kolodziejczyk (University of Warsaw, PL)*



Proving superpolynomial lower bounds for bounded-depth systems with a parity connective is one of the most famous long-standing open problems in proof complexity. I will review some known results about bounded-depth Frege with parity and its subsystems. In the process, I will try to motivate a few open problems in the area.

## 3.11 Automatizability

*Massimo Lauria (Sapienza University of Rome, IT)*



We give a tutorial on the concept of automatizability of proof systems, i.e. the possibility of finding relatively short proof efficiently. We survey known results and sketch the proof that resolution is not automatizable, by [1]. We conclude by surveying the results about the closely related concept of weak automatizability, and by discussing its connections with interpolation.

### References

1 Michael Alekhnovich, Alexander A. Razborov *Resolution is Not Automatizable Unless $W[P]$ is Tractable* FOCS 2001: 210–219

## 3.12 Are Short Proofs Narrow? QBF Resolution Is Not so Simple

*Meena Mahajan (Institute of Mathematical Sciences – Chennai, IN)*


**Joint work of** Olaf Beyersdorff, Leroy Chew, Meena Mahajan, Anil Shukla
**Main reference** Olaf Beyersdorff, Leroy Chew, Meena Mahajan, Anil Shukla: "Are Short Proofs Narrow? QBF Resolution Is *Not* So Simple", ACM Trans. Comput. Log., Vol. 19(1), pp. 1:1–1:26, 2018.
**URL** http://dx.doi.org/10.1145/3157053

One of the main techniques for proving size and space lower bounds in classical resolution proceeds via width: the results of Ben-Sasson and Wigderson [1] and of Atserias and Dalmau [2] show that lower bounds on width imply lower bounds on size and space respectively. We assess the effectiveness of such a technique for the QBF system QRes (used to prove QBFs false). Along the way, we show that the QBF proof systems Forall-Expansion+Resolution and IR-calc, provably separated in general, have the same power in their tree-like versions.

### References

1 Eli Ben-Sasson, Avi Wigderson: *Short proofs are narrow – resolution made simple.* J. ACM 48(2): 149–169 (2001)
2 Albert Atserias, Víctor Dalmau: *A combinatorial characterization of resolution width.* J. Comput. Syst. Sci. 74(3): 323–334 (2008)

## 3.13 Partially Definable Forcing

*Moritz Müller (Universität Wien, AT)*



The talk explains a general method of forcing to construct models of weak arithmetics relevant for propositional proof complexity. Proofs of independence results of Paris-Wilkie, Riis and Ajtai are naturally embedded in this framework.

## 3.14 Lower Bound Techniques for Nullstellensatz and Polynomial Calculus

*Jakob Nordström (KTH Royal Institute of Technology – Stockholm, SE)*



This talk is intended to give a high-level survey of techniques for proving lower bounds for Nullstellensatz and polynomial calculus. In particular, we will focus on the method in [1] for obtaining degree lower bounds in polynomial calculus using pseudo-ideals and pseudo-reductions, and on some further extensions presented in [2].

### References

1 Michael Alekhnovich, Alexander Razborov: *Lower Bounds for Polynomial Calculus: Non-Binomial Case*. Proceedings of the Steklov Institute of Mathematics 242: 18-35 (2003)
2 Mladen Miksa, Jakob Nordström: *A Generalized Method for Proving Polynomial Calculus Degree Lower Bounds*. Conference on Computational Complexity 2015: 467-487

## 3.15 Supercritical Space-Width Trade-offs for Resolution

*Jakob Nordström (KTH Royal Institute of Technology – Stockholm, SE)*


**Joint work of** Christoph Berkholz, Jakob Nordström
**Main reference** Christoph Berkholz, Jakob Nordström: "Supercritical Space-Width Trade-Offs for Resolution", in Proc. of the 43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy, LIPIcs, Vol. 55, pp. 57:1–57:14, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2016.
**URL** https://doi.org/10.4230/LIPIcs.ICALP.2016.57

We show that there are CNF formulas which can be refuted in resolution in both small space and small width, but for which any small-width resolution proof must have space exceeding by far the linear worst-case upper bound. This significantly strengthens the space-width trade-offs in [1], and provides one more example of trade-offs in the "supercritical" regime above worst case recently identified by [2]. We obtain our results by using Razborov's new hardness condensation technique and combining it with the space lower bounds in [3].

(This talk should have been given by Christoph Berkholz, who unfortunately had to cancel his participation on short notice.)

### References

1 Eli Ben-Sasson: *Size space tradeoffs for resolution*. SIAM Journal on Computing 28(6): 2511–2525 (2009)
2 Alexander A. Razborov: *A New Kind of Tradeoffs in Propositional Proof Complexity*. J. ACM 63(2): 16:1–16:14 (2016)
3 Eli Ben-Sasson, Jakob Nordström: *Short Proofs May Be Spacious: An Optimal Separation of Space and Length in Resolution*. FOCS 2008: 709–718

## 3.16 Sum-of-Squares, Counting Logics and Graph Isomorphism

*Joanna Ochremiak (University Paris-Diderot, FR)*


**Joint work of** Albert Atserias, Joanna Ochremiak
**Main reference** Albert Atserias, Joanna Ochremiak: "Definable Ellipsoid Method, Sums-of-Squares Proofs, and the Isomorphism Problem", CoRR, Vol. abs/1802.02388, 2018.
**URL** http://arxiv.org/abs/1802.02388

I will discuss recent joint work with Albert Atserias on connections between equivalence in finite variable logics with counting and semidefinite relaxations of the graph isomorphism problem.

## 3.17 Provability of Weak Circuit Lower Bounds

*Jan Pich (Universität Wien, AT)*


**Joint work of** Moritz Müller, Jan Pich
**Main reference** Moritz Müller, Ján Pich: "Feasibly constructive proofs of succinct weak circuit lower bounds", Electronic Colloquium on Computational Complexity (ECCC), Vol. 24, p. 144, 2017.
**URL** https://eccc.weizmann.ac.il/report/2017/144

The existing circuit lower bounds for explicit Boolean functions are very constructive, as captured in the notion of natural proofs. Following initial work of Razborov and Krajíček [1, 2], we investigate the constructive aspects of circuit lower bounds from the perspective of mathematical logic and show that $\mathrm{AC}^0$, $\mathrm{AC}^0[p]$ and monotone circuit lower bounds expressed by $\forall\Sigma^b_1$ formulas are provable in Jerabek's theory of approximate counting $\mathrm{APC}_1$. Consequently, we obtain short proofs of poly$(n)$-size tautologies expressing these circuit lower bounds, where n is the number of inputs of the circuit. These proofs take place in a slight extension of Extended Frege system. In case of Razborov-Smolensky's lower bound, we give a succinct version of natural proofs against $\mathrm{AC}^0[p]$ with proofs in a propositional proof system known as WF.

### References

1 Alexander Razborov: *Bounded arithmetic and lower bounds in Boolean complexity.* Feasible Mathematics II, 344–386 (1995)
2 Jan Krajíček: *Bounded arithmetic, propositional logic, and complexity theory.* Cambridge University Press, 1995.

## 3.18 Sum of Squares Lower Bounds from Symmetry and a Good Story

*Aaron Potechin (KTH Royal Institute of Technology – Stockholm, SE)*


**Main reference** Aaron Potechin: "Sum of squares lower bounds from symmetry and a good story", CoRR, Vol. abs/1711.11469, 2017.
**URL** http://arxiv.org/abs/1711.11469

The sum of squares hierarchy is a hierarchy of semidefinite programs which has the three advantages of being broadly applicable (it can be applied whenever the problem can be

phrased in terms of polynomial equations over $\mathbb{R}$), powerful (it captures the best known algorithms for several problems including max cut, sparsest cut, and unique games), and in some sense, simple (all it is really using is the fact that squares are non-negative over $\mathbb{R}$). The sum of squares hierarchy can also be viewed as the Positivstellensatz proof system.

## 3.19 Lifting Nullstellensatz Degree to Monotone Span Program Size

*Robert Robere (University of Toronto, CA)*


**Joint work of** Toniann Pitassi, Robert Robere
**Main reference** Toniann Pitassi, Robert Robere: "Lifting nullstellensatz to monotone span programs over any field", in Proc. of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018, pp. 1207–1219, ACM, 2018.
**URL** http://dx.doi.org/10.1145/3188745.3188914

Karchmer and Wigderson introduced an elegant model of computation, called span programs, which capture the complexity of computing with linear algebra over a field $\mathbb{F}$. In this talk, we discuss some recent work in which we characterize the monotone span program size of certain "structured" boolean functions in terms of Nullstellensatz degree over any field. This characterization leads to the resolution of a number of open problems on the complexity of span programs, including

- A superpolynomial separation between non-monotone span programs and span programs over characteristic 2,
- An exponential separation between monotone span programs over any field and monotone circuits, and
- A strongly exponential separation between monotone span programs over fields with different characteristic.

## 3.20 Monotone Circuit Lower Bounds from Resolution

*Dmitry Sokolov (KTH Royal Institute of Technology – Stockholm, SE)*


**Joint work of** Ankit Garg, Mika Göös, Pritish Kamath, Dmitry Sokolov
**Main reference** Ankit Garg, Mika Göös, Pritish Kamath, Dmitry Sokolov: "Monotone circuit lower bounds from resolution", in Proc. of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018, pp. 902–911, ACM, 2018.
**URL** http://dx.doi.org/10.1145/3188745.3188838

For any unsatisfiable CNF formula $F$ that is hard to refute in the resolution proof system, we show that a gadget-composed version of $F$ is hard to refute in any proof system whose lines are computed by efficient communication protocols (in particular, as in cutting planes)—or, equivalently, that a monotone function associated with $F$ has large monotone circuit complexity.

This result is essentially a lifting theorem for "decision dags" and "dag communication protocols."

## 3.21 Bounded Arithmetic and Propositional Upper Bounds

*Neil Thapen (The Czech Academy of Sciences – Prague, CZ)*



I will talk about how bounded arithmetic can be used to prove, and understand, propositional upper bounds. I will briefly survey some results of this kind, and then talk in some detail about an example, a simple first-order theory that captures the kind of reasoning you can do in resolution. In particular, if you can prove something in the theory, then you get polynomial size resolution refutations. The other direction also holds, modulo some issues of uniformity, and the construction generalizes to other fragments of $AC^0$-Frege.

## 3.22 Bounded Arithmetic Does Not Collapse to Approximate Counting

*Neil Thapen (The Czech Academy of Sciences – Prague, CZ)*


**Joint work of** Leszek Kolodziejczyk; Neil Thapen

We adapt the "fixing lemma", a simple switching lemma used recently to show lower bounds for random resolution, to show that Jerabek's theory of approximate counting does not prove the CPLS principle (coloured polynomial local search). This settles an open problem by showing that bounded arithmetic is strictly stronger than approximate counting, if we compare the strength of theories by looking at their $\forall\Sigma^{b}_{1}$ consequences.

## 3.23 Cops-Robber games and the resolution of Tseitin formulas

*Jacobo Torán (Universität Ulm, DE)*


**Joint work of** Nicola Galesi, Navid Talebanfard, Jacobo Torán
**Main reference** Nicola Galesi, Navid Talebanfard, Jacobo Torán: "Cops-Robber Games and the Resolution of Tseitin Formulas", in Proc. of the Theory and Applications of Satisfiability Testing – SAT 2018 – 21st International Conference, SAT 2018, Held as Part of the Federated Logic Conference, FloC 2018, Oxford, UK, July 9-12, 2018, Proceedings, Lecture Notes in Computer Science, Vol. 10929, pp. 311–326, Springer, 2018.
**URL** http://dx.doi.org/10.1007/978-3-319-94144-8_19

We characterize several complexity measures for the resolution of Tseitin formulas in terms of a two person cop-robber game. Our game is a slight variation of the the one Seymour and Thomas used in order to characterize the tree-width parameter. For any undirected graph, by counting the number of cops needed in our game in order to catch a robber in it, we are able to exactly characterize the width, variable space and depth measures for the resolution of the Tseitin formula corresponding to that graph. We also give an exact game characterization of resolution variable space for any formula.

We show that our game can be played in a monotonous way. This implies that the corresponding resolution measures on Tseitin formulas correspond to those under the restriction of regular resolution.

Using our characterizations we improve the existing complexity bounds for Tseitin formulas showing that resolution width, depth and variable space coincide up to a logarithmic factor, and that variable space is bounded by the clause space times a logarithmic factor.

## 3.24 Nullstellensatz is Polynomially Equivalent to Sum-of-Squares over Algebraic Circuits

*Iddo Tzameret (Royal Holloway, University of London, GB)*


**Joint work of** Edward Hirsch, Iddo Tzameret

We consider the relative strength of algebraic and semi-algebraic proof systems when the complexity of proofs is measured by algebraic circuit size (in contrast to degree). We show that under this measure, Nullstellensatz simulates Sum-of-Squares proofs and Sherali-Adams. This contrasts known separations between the Nullstellensatz and Sum-of-Squares in the degree regime.

## 3.25 Proof Systems for Pseudo-Boolean SAT Solving

*Marc Vinyals (TIFR Mumbai, IN)*


**Joint work of** Marc Vinyals, Jan Elffers, Jesús Giráldez-Cru, Stephan Gocht, Jakob Nordström
**Main reference** Marc Vinyals, Jan Elffers, Jesús Giráldez-Cru, Stephan Gocht, Jakob Nordström: "In Between Resolution and Cutting Planes: A Study of Proof Systems for Pseudo-Boolean SAT Solving", in Proc. of the Theory and Applications of Satisfiability Testing – SAT 2018 – 21st International Conference, SAT 2018, Held as Part of the Federated Logic Conference, FloC 2018, Oxford, UK, July 9-12, 2018, Proceedings, Lecture Notes in Computer Science, Vol. 10929, pp. 292–310, Springer, 2018.
**URL** http://dx.doi.org/10.1007/978-3-319-94144-8_18

Current SAT solvers reason within the resolution proof system, and that gives them a big advantage with respect to DPLL solvers, which are limited to tree-like resolution. Pseudo-Boolean solvers can reason within cutting planes, hence they are potentially more powerful, but implementation constraints dictate that they are limited to a subset of inference rules. A natural question, then, is whether these rules are enough to exploit the full power of cutting planes.

In this talk we identify subsystems of cutting planes that arise from these limited rules and we classify them, showing in particular that pseudo-Boolean solvers are limited to resolution if their input is encoded adversarially. Additionally we craft formulas that we conjecture able to separate these proof systems at a more fundamental level.

# 4 A List of Some Open Problems

Below follows a (non-exhaustive) list of open research problems discussed during the seminar week. We have collected them in this report in the hope that this can serve as a convenient point of reference for the community, and in the longer term perhaps inspire the collection of open problems in proof complexity in a community research wiki or similar.

## 4.1 Simulation/Separation of Semi-algebraic Proof Systems

*Paul Beame (University of Washington – Seattle, WA, beame@cs.washington.edu)*



### 4.1.1 Preliminaries

I will assume familiarity with semi-algebraic proof systems: Cutting Planes, LS, LS+, Sherali-Adams, and SOS proof systems, as well as Tseitin tautologies.

### 4.1.2 Problems

With the exception of recent work on extension complexity lower bounds, much of the discussion of semi-algebraic proof systems is focused on rank (or degree) and not on proof size.

▶ **Open Problem 1.** *Can LS, LS+, or SOS proofs p-simulate Cutting Planes proofs for translations of Boolean formulas?*

Buss and Clote [1] showed that Cutting Planes proofs are polynomially equivalent to a restricted form of such proofs in which the division rule is only applied with divisor 2. One natural family of Boolean formulas that use this inference consists of the Tseitin formulas on bounded-degree graphs. Another particularly natural graph property to consider is the matching principle on $K_{2n+1}$ which is known as the Parity Principle: "There is no perfect matching on $K_{2n+1}$". This is expressed as the following system of inequalities which is a direct translation of the clausal forms:

$$\begin{aligned}
\sum_{i\in e} x_e &\geq 1 \qquad 1 \leq i \leq 2n+1,\ e \in \binom{[2n+1]}{2} \\
x_e + x_f &\leq 1 \qquad e, f \in \binom{[2n+1]}{2},\ e \cap f \neq \emptyset \\
x_e &\geq 0 \qquad e \in \binom{[2n+1]}{2} \\
x_e &\leq 1 \qquad e \in \binom{[2n+1]}{2}
\end{aligned}$$

It is easy for all of the semi-algebraic proof systems above to derive

$$\sum_{i\in e} x_e \leq 1 \qquad 1 \leq i \leq 2n+1,\ e \in \binom{[2n+1]}{2}$$

in small size. Then by adding these inequalities one obtains:

$$2 \sum_{e\in\binom{[2n+1]}{2}} x_e \leq 2n+1$$

In Cutting Planes with divisor 2 one can now round this to obtain:

$$\sum_{e \in \binom{[2n+1]}{2}} x_e \leq n$$

and using this one easily obtains a contradiction in any of the systems. The only hard part is the division rule. Therefore it is natural to ask:

▶ **Open Problem 2.** *Are there polynomial-size LS, LS+, or SOS proofs of the Parity Principle?*

This was essentially asked by Lovasz at the 1996 Oberwolfach complexity theory workshop for the case of LS, LS+ by asking about proofs of stable set size bounds for a particular family of graphs, the line graphs of $K_{2n+1}$, which is an equivalent question to the one for the Parity Principle. It seems reasonable to conjecture that the answer to both of the above open problems is no.

Since the only hard part of this inference is the one line of division by 2, Open Problem 1 could be resolved depending on the outcome of the following:

▶ **Open Problem 3.** *For what values of $m$ and $n$ do LS, LS+, or SOS proofs have polynomial-size proofs of the following of inference?*

*Given*

$$\begin{aligned} 2\sum_{i=1}^{n} x_i &\leq 2m+1, \\ x_i &\geq 0 \qquad 1 \leq i \leq n \\ x_i &\leq 1 \qquad 1 \leq i \leq n \end{aligned}$$

*infer*

$$\sum_{i=1}^{n} x_i \leq m$$

Note that Grigoriev's work [2] on Postivstellensatz (SOS) proofs of the above constraints, which he calls the knapsack inequalities, yields large rank lower bounds for the case that $m$ is near $n/2$ (within roughly $\pm\sqrt{n}$). By the extension complexity results of Lee, Raghavendra, and Steurer [3] this implies exponential size lower bounds in this case. In the case of the Parity Principle, $m$ is $\Theta(\sqrt{n})$ so it is not covered by that bound.

## References


1 Samuel R. Buss and Peter Clote. Cutting planes, connectivity, and threshold logic. *Arch. Math. Log.*, 35(1):33–62, 1996.

2 Dima Grigoriev. Complexity of positivstellensatz proofs for the knapsack. *Computational Complexity*, 10(2):139–154, 2001

3 James R. Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 567–576, 2015.

## 4.2 Geometric Lower Bounds for Cutting Planes

*Yuval Filmus (Technion – Haifa, IL, yuvalfi@cs.technion.ac.il)*



Come up with a lower bound technique for cutting planes that, as opposed to the interpolation method or DAG-like communication, does not a reduction to circuit complexity. For example, a geometric method based on properties of polytopes, like algebraic decision tree lower bounds.

## 4.3 The Effect of Arity on the Power of Semantic Cutting Planes

*Yuval Filmus (Technion – Haifa, IL, yuvalfi@cs.technion.ac.il)*



### 4.3.1 Preliminaries

Cutting planes is usually defined with the syntactic rules *addition* and *division*. The first rule allows deducing from $\sum_i a'_i x_i \geq b'$ and $\sum_i a''_i x_i \geq b''$ the line $\sum_i (c'a'_i + c''a''_i)x_i \geq c'b' + c''b''$ for all integers $c', c'' \geq 0$, and the second rule allows deducing from $\sum_i ca_i x_i \geq b$ the line $\sum_i a_i x_i \geq \lceil b/c \rceil$ for all $c \geq 1$.

One can augment these rules with semantic rules. The proof system *k-ary semantic cutting planes* allows deducing a line $L$ from lines $L_1, \ldots, L_k$ as long as every $0,1$-assignment which satisfies $L_1, \ldots, L_k$ also satisfies $L$. Note that when $k = 2$, the syntactic rules are no longer necessary, and that when $k = 1$, we only need the syntactic rule of addition.

Filmus, Hrubeš and Lauria [1] showed that unary semantic cutting planes cannot be p-simulated by syntactic cutting planes, and proved exponential lower bounds on $n^\epsilon$-ary semantic cutting planes.

### 4.3.2 Problem

▶ **Open Problem 4.** *Let $1 \leq k_1 < k_2$ be constants. Does $k_1$-ary semantic cutting planes p-simulate $k_2$-ary semantic cutting planes?*

Hrubeš and Pudlák [2] gave an affirmative answer for the analogous question on monotone real circuits.

### References

1 Yuval Filmus, Pavel Hrubeš, Massimo Lauria: *Semantic versus Syntactic Cutting Planes.* STACS 2016: 35:1–35:13

2 Pavel Hrubeš, Pavel Pudlák *A note on monotone circuits.* Inf. Process. Lett. 131: 15–19 (2018)

## 4.4 Questions on Ideal Proof Systems

*Joshua A. Grochow (University of Colorado – Boulder, USA, jgrochow@colorado.edu)*



### 4.4.1 Preliminaries

▶ **Definition 1** (Ideal Proof System [4, Def. 1.9] (cf. [5, 6]))**.** An *IPS certificate* that a polynomial $G(\vec{x}) \in \mathbb{F}[\vec{x}]$ is in the ideal [respectively, radical of the ideal] generated by $F_1(\vec{x}), \ldots, F_m(\vec{x})$ is a polynomial $C(\vec{x}, \vec{y})$ such that
1. $C(\vec{x}, \vec{0}) = 0$, and
2. $C(\vec{x}, F_1(\vec{x}), \ldots, F_m(\vec{x})) = G(\vec{x})$ [respectively, $G(\vec{x})^k$ for any $k > 0$].

An *IPS derivation* of $G$ [resp. $G^k$] from $F_1, \ldots, F_m$ is a circuit computing some IPS certificate that $G \in \langle F_1, \ldots, F_m \rangle$ [resp., $G \in \sqrt{\langle F_1, \ldots, F_m \rangle}$].

When applied as a proof system of unsatisfiability of Boolean formulas, we translate a CNF $\varphi$ into a system of equations as follows, and an IPS proof is a derivation that 1 is in the ideal generated by the following polynomials. We translate a clause $\kappa$ of $\varphi$ into a single algebraic equation $F(\vec{x})$ as follows: $x \mapsto 1 - x$, $x \vee y \mapsto xy$. This translation has the property that a $\{0,1\}$ assignment satisfies $\kappa$ if and only if it satisfies the equation $F = 0$. Let $\kappa_1, \ldots, \kappa_m$ denote all the clauses of $\varphi$, and let $F_i$ be the corresponding polynomials. Then the system of equations we consider is $F_1(\vec{x}) = \cdots = F_m(\vec{x}) = x_1^2 - x_1 = \cdots = x_n^2 - x_n = 0$. The latter equations force any solution to this system of equations to be $\{0,1\}$-valued. (Note that, in principle, Boolean tautologies can be refuted without the Boolean axioms $x_i^2 - x_i$, but we do not know how this affects the complexity of the proofs in general.)

To motivate the following variant of IPS, we may consider

$$F_1(x_1, \ldots, x_n), \ldots, F_m(x_1, \ldots, x_n)$$

as a polynomial map $F = (F_1, \ldots, F_m) \colon \mathbb{F}^n \to \mathbb{F}^m$. Then this system of polynomials has a common zero if and only if $\vec{0}$ is the image of $F$. In fact, Grochow and Pitassi [4, Appendix B] show that for any system of equations coming from an unsatisfiable Boolean CNF, the system of polynomials has a common zero if and only if $\vec{0}$ is in the *closure* of the image of $F$. This holds regardless of whether the equations include $x_i^2 - x_i = 0$, $x_i^2 - 1 = 0$, or neither of these, though at the moment the proof only works over algebraically closed fields and over dense subfields of $\mathbb{C}$ (such as $\mathbb{Q}(i)$).

▶ **Definition 2** (The Geometric Ideal Proof System [4, App. B])**.** A *geometric IPS certificate* that a system of $\mathbb{F}$-polynomial equations $F_1(\vec{x}) = \cdots = F_m(\vec{x}) = 0$ is unsatisfiable over $\overline{\mathbb{F}}$ is a polynomial $C \in \mathbb{F}[y_1, \ldots, y_m]$ such that
1. $C(0, 0, \ldots, 0) = 1$, and
2. $C(F_1(\vec{x}), \ldots, F_m(\vec{x})) = 0$. In other words, $C$ is a polynomial relation amongst the $F_i$.

A *geometric IPS proof* of the unsatisfiability of $F_1 = \cdots = F_m = 0$, or a *geometric IPS refutation* of $F_1 = \cdots = F_m = 0$, is an $\mathbb{F}$-algebraic circuit on inputs $y_1, \ldots, y_m$ computing some geometric certificate of unsatisfiability.

If $C$ is a geometric certificate, then $1 - C$ is an IPS certificate that involves only the $y_i$. Hence the smallest circuit size of any geometric certificate is at least the smallest circuit size of any algebraic certificate. We do not know, however, if these complexity measures are polynomially related, as highlighted in a question below.

We call a system of equations "standard Boolean" if it includes $x_i^2 = x_i$ for all $i$, and "multiplicative Boolean" if it includes $x_i^2 = 1$ for all $i$; by "Boolean system of equations" we mean either of these.

### 4.4.2 Problems

▶ **Open Problem 5** (Hrubeš [7]). *Find a polynomial $f$ that vanishes on $\{0,1\}^n$ such that any IPS certificate showing that $f \in \langle x_i^2 - x_i | x \in [n] \rangle$ requires super-polynomial algebraic circuit size.*

Of course, if the $f$ is the translation of an unsatisfiable Boolean CNF, then its existence would imply VP $\neq$ VNP, and moreover such a CNF-translation $f$ must exist assuming NP $\not\subseteq$ coAM. A conditional result would also be interesting here, so long as the condition is weaker than NP $\not\subseteq$ coAM; perhaps the most interesting would be finding such an $f$ assuming only VP $\neq$ VNP.

▶ **Open Problem 6** ([4, Open Question 8.2]). *Let $\beta \notin \{0, \ldots, 2n\}$, and let $\mathbb{F}$ be a field of characteristic at least $2n+1$. Prove lower bounds on restricted versions of IPS certificates (as in, e. g., [1]) for the unsatisfiable system of equations*

$$x_1 + \cdots + x_n - x = x_{n+1} + \cdots + x_{2n} - x' = x + x' - \beta = x_1^2 - x_1 = \cdots = x_n^2 - x_n = 0.$$

▶ **Open Problem 7** ([4, Open Question A.12]). *Does every IPS certificate for the $n \times n$ Inversion Principle $XY = I \Rightarrow YX = I$ require computing a determinant? That is, is it the case that for every IPS certificate $C$, some determinant of size $n^{\Omega(1)}$ reduces to $C$ by a $O(\log n)$-depth circuit reduction?*

A positive answer here would show that, indeed, the Inversion Principle does not have an IPS proof of logarithmic depth unless the determinant has polynomial-size algebraic formulas.

▶ **Open Problem 8** ([4, Open Question B.4]). *For Boolean systems of equations, is Geometric IPS polynomially equivalent to IPS? That is, is there always a geometric certificate whose circuit size is at most a polynomial in the circuit size of the smallest algebraic certificate?*

For radical membership, an exponential degree upper bound is known (often called Effective Nullstellensatz), and known to be tight, but we could ask about strengthening such bounds to circuit size. For ideal membership, we observed that a subexponential IPS size upper bound would violate the Space Hierarchy Theorem because ideal membership in general is EXPSPACE-complete. But for radical membership, we do not know how to rule this out.

▶ **Open Problem 9** ([4, Open Question 1.11]). *For any*

$$G(\vec{x}) \in \sqrt{\langle F_1(\vec{x}), \ldots, F_m(\vec{x}) \rangle}$$

*is there always an IPS-certificate of subexponential size that $G$ is in the* radical *of $\langle F_1, \ldots, F_m \rangle$? Similarly, for $G, F_1, \ldots, F_m \in \mathbb{Z}[x_1, \ldots, x_n]$, is there a constant-free $\mathrm{IPS}_{\mathbb{Z}}$-certificate of subexponential size that $aG(\vec{x})$ is in the* radical *of the ideal $\langle F_1, \ldots, F_m \rangle$ for some integer $a$?*

▶ **Open Problem 10** ([4, General Question 7.4]). *Given a family of cosets of ideals $f_n^{(0)} + I_n$ (or more generally modules) of polynomials, with $I_n \subseteq R[x_1, \ldots, x_{poly(n)}]$, consider the function families $(f_n) \in (f_n^{(0)} + I_n)$ (meaning that $f_n \in f_n^{(0)} + I_n$ for all $n$) under any computational reducibility $\leq$ such as p-projections. What can the $\leq$ structure look like? For example:*

a. *When, if ever, is there such a unique $\leq$-minimum (even a single nontrivial example would be interesting)?*
b. *Can there be infinitely many incomparable $\leq$-minima?*
c. *Say a $\leq$-degree $\mathbf{d}$ is "saturated" in $(f_n^{(0)} + I_n)$ if every $\leq$-degree $\mathbf{d}' \geq \mathbf{d}$ has some representative in $f^{(0)} + I$. Must saturated degrees always exist? We suspect yes, given that one may multiply any element of $I$ by arbitrarily complex polynomials.*
d. *What can the set of saturated degrees look like for a given $(f_n^{(0)} + I_n)$?*
e. *Must every $\leq$-degree in $f^{(0)} + I$ be* below *some saturated degree?*
f. *What can the $\leq$-structure of $f^{(0)} + I$ look like below a saturated degree?*
g. *...*

Problem 10 is of interest even when $f^{(0)} = 0$, that is, for ideals and modules of functions rather than their nontrivial cosets. For ideals, these questions are also related to algebraic natural proofs [2, 3].

### References


1 Michael A. Forbes, Amir Shpilka, Iddo Tzameret, Avi Wigderson: *Proof Complexity Lower Bounds from Algebraic Circuit Complexity.* CCC 2016: 32:1–32:17
2 Michael A. Forbes, Amir Shpilka, Ben Lee Volk: *Succinct hitting sets and barriers to proving algebraic circuits lower bounds.* STOC 2017: 653–664
3 Joshua A. Grochow, Mrinal Kumar, Michael Saks, Shubhangi Saraf: *Towards an algebraic natural proofs barrier via polynomial identity testing.* Electronic Colloquium on Computational Complexity (ECCC) 24: 009 (2017)
4 Joshua A. Grochow, Toniann Pitassi: *Circuit complexity, proof complexity, and polynomial identity testing.* FOCS 2014
5 Toniann Pitassi: *Algebraic propositional proof systems.* Descriptive Complexity and Finite Models 1996: 215–244
6 Toniann Pitassi: *Propositional proof complexity and unsolvability of polynomial equations.* ICM 1998: 215–244
7 Pavel Hrubeš: *Arithmetic circuits and proof complexity* Algebraic Complexity Theory 2016


## 4.5 The Complexity of Linear Resolution

*Jan Johannsen (Ludwig-Maximillians-Universität München, DE, jan.johannsen@ifi.lmu.de)*



### 4.5.1 Preliminaries

A linear resolution refutation of a CNF formula $F$ is a sequence of clauses $C_1, \ldots, C_m$ such that

- $C_1$ is a clause from $F$,
- $C_m$ is the empty clause, and
- each clause $C_{i+}$ is obtained by resolution from $C_i$ and either a clause $D$ from $F$, or an earlier clause $C_j$ for $j < i$.

In other words, a resolution refutation is linear if in every resolution step, one of the used clauses is the one derived in the immediately preceding step.

It is now known that linear resolution $p$-simulates tree-like resolution, but is not simulated by regular resolution [1].

### 4.5.2 Problem

The relationship between linear and full resolution with respect to $p$-simulation is a long-standing open problem.

▶ **Open Problem 11.** *Is there a super-polynomial or even exponential separation between linear and unrestricted resolution? Or does linear resolution p-simulate unrestricted resolution?*

#### References

1 S. R. Buss, J. Johannsen, On Linear Resolution, *Journal on Satisfiability, Boolean Modeling and Computation*, 16:23–35, 2017.

## 4.6 New Hard Examples for Regular Resolution

*Jan Johannsen (Ludwig-Maximillians-Universität München, DE, jan.johannsen@ifi.lmu.de)*



### 4.6.1 Preliminaries

A (dag-like) resolution refutation is *regular* if on every path through the proof dag every variable is resolved on at most once. There are several examples that witness an exponential separation of regular from unrestricted dag-like resolution [1, 4].

An ongoing direction of research tries to analyse the complexity of refinements of resolution that correspond to contemporary SAT algorithms using conflict-driven clause learning. These refinements are between regular and full dag-like resolution w.r.t. size complexity. There are polynomial upper bounds in these systems for all the hard examples mentioned above [2, 3], so they can have an exponential speed-up over regular resolution.

### 4.6.2 Problem

An open question is to give a super-polynomial or exponential separation between these clause learning proof systems and full resolution. Any separating example needs to necessarily also separate regular from full resolution. But for all such known exaples we have polynomial upper bounds. So to attack this problem, we first need to solve the following:

▶ **Open Problem 12.** *Find new examples of families of formulas that have polynomial size resolution refutations, but require exponential size regular resolution refutations.*

#### References

1 Michael Alekhnovich, Jan Johannsen, Toniann Pitassi, Alasdair Urquhart. An exponential separation between regular and general resolution. *Theory of Computing*, **3**(5):81–102, 2007.
2 Maria Luisa Bonet, Samuel R. Buss, Jan Johannsen. Improved separations of regular resolution from clause learning proof systems. *Journal of Artificial Intelligence Research*, 49:669–703, 2014.
3 S. Buss, L. Kołodziejczyk, *Small stone in pool. Logical Methods in Computer Science*, 10(2), 2014, paper 16.
4 A. Urquhart, *A near-optimal separation of regular and general resolution*, SIAM Journal on Computing, 40 (2011), pp. 107–121.

## 4.7 R(Lin/$\mathbb{F}_2$) Lower Bounds via Randomised Feasible Interpolation

*Igor C. Oliveira (University of Oxford, GB, igor.carboni.oliveira@cs.ox.ac.uk)*



### 4.7.1 Preliminaries

We are interested in the problem of establishing (dag-like) lower bounds for R(Lin/$\mathbb{F}_2$), a proof system that corresponds to resolution extended with linear equations over the field $\mathbb{F}_2$. For more details about this proof system, we refer to Itsykson and Sokolov [1], where *tree-like* lower bounds are also described. (Note that the work of Buss, Kolodziejczyk, and Zdanowski [2] shows a collapse of $F_d[\oplus]$-Frege to depth three, which further motivates the study of R(Lin/$\mathbb{F}_2$) and its extensions.)

More recently, Krajíček [4] proposed an extension of the feasible interpolation technique that employs randomized communication complexity, and that allows one to reduce lower bounds for R(Lin/$\mathbb{F}_2$) and other proof systems to the investigation of monotone circuits with local oracles. This is an extension of monotone circuits that incorporates extra inputs (local oracles) to help the computation. While super-polynomial lower bounds against monotone circuits with local oracles for computational problems such as clique vs. colorings would provide lower bounds for R(Lin/$\mathbb{F}_2$), currently only restricted lower bounds against such circuits are known [3].

We refer to the last paper for a precise definition of this circuit model. Here we only recall that a parameter $\mu$ measures the power of the local oracles. (It is connected to the failure probability of certain randomised communication protocols derived from propositional proofs.) This parameter appears in the statement of the problem, described next.

### 4.7.2 Problems

Let $k \geq 3$ be a positive integer, $U_{n,k}$ be the set of $n$-vertex graphs corresponding to $k$-cliques, and $V_{n,k}$ be the set of complete $(k-1)$-partite graphs over $n$ vertices. Show that any monotone circuit with local oracles and locality $\mu \leq 1/100$ that separates $U_{n,k}$ and $V_{n,k}$ must have super-polynomial size (say, for some super-constant function $k(n) \leq n$).

We are also interested in non-trivial results for $k = 3$ (triangles vs. complete bipartite graphs). While lower bounds in this regime will not have important consequences in proof complexity, they might shed light into the power and limitations of this circuit model, and further inform the randomised feasible interpolation program.

### References

1 Dmitry Itsykson, Dmitry Sokolov: *Lower Bounds for Splittings by Linear Combinations.* MFCS (2) 2014: 372–383
2 S. R. Buss, L. A. Kołodziejczyk, K. Zdanowski: *Collapsing modular counting in bounded arithmetic and constant depth propositional proofs.* Transactions of the AMS 367 (2015), 7517–7563.
3 Jan Krajíček, Igor Carboni Oliveira: *On monotone circuits with local oracles and clique lower bounds.* CoRR abs/1704.06241 (2017)
4 Jan Krajíček: *Randomized feasible interpolation and monotone circuits with a local oracle.* CoRR abs/1611.08680 (2016)

## 4.8 Unprovability of Circuit Upper Bounds in Logical Theories

*Igor C. Oliveira (University of Oxford, GB, igor.carboni.oliveira@cs.ox.ac.uk)*



### 4.8.1 Preliminaries

It is believed that $\mathsf{NP} \not\subseteq \mathsf{P/poly}$, but it is consistent with our knowledge that $\mathsf{NTIME}[2^n] \subseteq \mathsf{SIZE}[O(n)]$. Given the lack of techniques for proving non-trivial lower bounds, we are interested in the logical complexity/(un)provability aspects of circuit complexity theory. This research program is a few decades old, but for brevity we restrict our discussion to a small number of references more directly connected to our problem.

Cook's theory PV [1] or its mild extensions seem to formalize a large fraction of contemporary complexity theory. (We refer to the recent work of Muller and Pich [2] for more background on the formalization of circuit complexity in bounded arithmetic.) It is therefore of interest to understand when a given conjecture is provable or at least consistent with PV. We believe that NP requires large circuits, but since we don't know how to establish this result at this point, can we at least show that PV does *not* prove that $\mathsf{NP} \subseteq \mathsf{SIZE}[100n]$?

Cook and Krajíček [3] established conditional results of this form for PV and $S^1_2$. More recently, Krajíček and Oliveira [4] unconditionally showed that PV does not prove that P (polynomial time) is contained in $\mathsf{SIZE}[n^k]$, when $k$ is a fixed constant. In particular, there is a model $\mathfrak{M}$ of PV where a lot of complexity theory holds, and moreover in $\mathfrak{M}$ there are languages in P that cannot be computed by circuits of size $n^{100}$.

We would like to extend this theorem to an unprovability result for stronger logical theories. A natural candidate is the theory APC1 investigated by E. Jerabek and other authors. This theory extends PV and allows the formalization of many probabilistic constructions and randomised algorithms. Formally, APC1 adds to the axioms of PV a dual weak pigeonhole principle for polynomial-time function symbols. With enough work, this can be used to (approximately) formalize probabilities and events. We refer to Jerabek's related work and Muller and Pich [2] for further details.

### 4.8.2 Problem

Let $\mathsf{UP}_{k,c}(f)$ be the upper bound sentence (in the language of PV) from Krajíček and Oliveira [4] stating that the language encoded by the function symbol $f$ can be computed by circuits of size at most $c \cdot n^k$. Show that for each $k \geq 1$ there is a function symbol $g$ in the language of PV such that for no constant $c \geq 1$ APC1 proves the sentence $\mathsf{UP}_{k,c}(g)$.

We believe that a solution to this problem will require interesting new ideas from logic and complexity theory.

### References

1 Stephen Cook: *Feasibly constructive proofs and the propositional calculus.* STOC 1975, 83–97.
2 Moritz Müller, Ján Pich: *Feasibly constructive proofs of succinct weak circuit lower bounds.* Electronic Colloquium on Computational Complexity (ECCC) 24: 144 (2017)
3 Stephen A. Cook, Jan Krajíček: Consequences of the provability of NP ⊆ P/poly. J. Symb. Log. 72(4): 1353–1371 (2007)
4 Jan Krajíček, Igor Carboni Oliveira: *Unprovability of circuit upper bounds in Cook's theory PV.* Logical Methods in Computer Science 13(1) (2017)

## 4.9 Dag Communication Lower Bounds

*Dmitry Sokolov (KTH Royal Institute of Technology – Stockholm, SE, sokolovd@kth.se)*



▶ **Definition 1.** Let $U, V \in \{0,1\}^n$ be two sets. Let us consider a triple $(H, A, B)$, where $H$ is a directed acyclic graph, $A : H \times U \to \mathbb{N}$ and $B : H \times V \to \mathbb{N}$. We say that vertex $h \in H$ is valid for pair $(x, y) \in U \times V$ iff $A(h, x) = B(h, y) = 1$. We call this triple a *EQ dag protocol* for the pair $(U, V)$ and some relation $N : U \times V \to T$, where $T$ is a finite set of "possible answers", if the following holds:

- $H$ is an acyclic graph and the out-degree of all its vertices is at most 2;
- the leaves of $H$ are marked by element of $T$;
- there is a *root* $s \in H$ with in-degree 0 and this vertex is valid for all pairs from $U \times V$;
- if $h \in H$ is valid for pair $(x, y)$ and $h$ is not a leaf then at least one child of $h$ is valid for $(x, y)$;
- if $h \in H$ is valid for pair $(x, y)$, $h$ is a leaf and $h$ is marked by $t \in T$ then $t \in N(x, y)$.

The size of the game is the size of the graph $H$.

We say that we have *boolean dag protocol* iff vertex is valid in case that $A(h, x) = B(h, y) = 1$.

▶ **Definition 2.** Canonical search problem $Search_\varphi$ for an unsatisfiable formula $\varphi(x, y)$ in CNF: Alice receives values for the variables $x$, Bob receives values for the variables $y$, and their goal is to find a clause of $\varphi$ such that it is unsatisfied by this substitution.

We know that in case of boolean protocols an analog of Karchmer–Wigderson Theorem holds for boolean protocols (for $KW$ and $KW^m$ relations) and (monotone) circuits. If we apply this protocols for canonical search problem this protocols capture the huge class of proof systems. And we can prove lower bound for boolean protocols.

▶ **Open Problem 13.** *Can one prove lower bounds on EQ dag protocols for $Search_\varphi$ or $KW^m$ relations?*

▶ **Open Problem 14.** *In boolean case can we prove lower bound for three players in NOF model for $Search_{\varphi(x,y,z)}$ relation (vertex is valid iff $A(h, x, y) = B(h, y, z) = C(h, x, z) = 1$)?*

## 4.10 Game Characterization of Resolution Space

*Jacobo Torán (University of Ulm, DE, jacobo.toran@uni-ulm.de)*



### 4.10.1 Preliminaries

Game characterizations of complexity measures in resolution have helped to better understand these measures and the relations among them. Such game characterizations exist for width [1], space in tree-like resolution [2], depth [3] and variable space [4].

### 4.10.2 Problem

Is there a characterization of resolution clause space in terms of a combinatorial game?

**References**

1 Albert Atserias, Víctor Dalmau: *A Combinatorial Characterization of Resolution Width.* CCC 2003: 239–247
2 Juan Luis Esteban, Jacobo Torán: *A combinatorial characterization of treelike resolution space.* Inf. Process. Lett. 87(6): 295–300 (2003)
3 Alasdair Urquhart: *The Depth of Resolution Proofs.* Studia Logica 99(1-3): 349–364 (2011)
4 Nicola Galesi, Navid Talebanfard and Jacobo Torán: *Cops-Robber games and the resolution of Tseitin formulas.* SAT 2018

## 4.11 Miters

*Alasdair Urquhart (University of Toronto – Toronto, CA, urquhart@cs.toronto.edu)*



### 4.11.1 Preliminaries

A "miter" is a type of problem considered by hardware designers. Given a circuit $C$, with inputs $x_1, \ldots, x_n$, and gates $g_1, \ldots, g_m$, construct an isomorphic circuit $C'$ with gates $g'_1, \ldots, g'_m$. The miter $M(C)$ is the CNF formula formalizing the statement "$C$ and $C'$ give different outputs for the inputs $x_1, \ldots, x_n$."

Obviously, this statement is unsatisfiable, and what is more, it has a short, narrow resolution refutation. However, CDCL solvers have a hard time with such statements. Donald Knuth [1] describes this situation as "somewhat scandalous."

### 4.11.2 Problem

The problem is simply to give a good theoretical explanation of what is going on here.

**References**

1 Donald Knuth *The Art of Computer Programming, Volume 4, Fascicle 6, "Satisfiability", p. 121*

# 5 Examples of Outcomes of the Workshop

It still a bit too early for any concrete publications to have resulted from the workshop, but participants have reported that the the following papers, in different stages of preparation, were significantly influenced by discussions during the workshop:

**References**

1 Olaf Beyersdorff, Leroy Chew, Judith Clymo and Meena Mahajan: *Short Proofs in QBF Expansion.* Submitted
2 Stefan Dantchev, Nicola Galesi and Barnaby Martin: *Resolution and the binary encoding of combinatorial principles.* Manuscript in preparation
3 Jan Elffers, Jesús Giráldez-Cru, Jakob Nordström, and Marc Vinyals: *Using Combinatorial Benchmarks to Probe the Reasoning Power of Pseudo-Boolean Solvers.* SAT 2018
4 Nicola Galesi, Navid Talebanfard and Jacobo Torán: *Cops-Robber games and the resolution of Tseitin formulas.* SAT 2018

5 Alasdair Urquhart: *Switching lemmas and bounded depth Frege proofs.* Manuscript in preparation

Participants of the workshop have reported about other concrete research projects that resulted to a large part from contacts during the week at Dagstuhl. Since many of these projects are still in a start-up phase it would seem slightly premature to list concrete participants, but it can be mentioned that these projects involve researchers from the Academy of Sciences of the Czech Republic, KTH Royal Institute of Technology, Ludwig Maximilians Universität München, Tata Institute of Fundamental Research, University of Toronto, and University of Warsaw, in various constellations.

## 6 Evaluation by Participants

In addition to the traditional Dagstuhl evaluation after the workshop, the organizing committee also arranged for a separate evaluation which specific questions about different aspects of the workshop. Below follows a summary of the answers.

The participants unanimously praise three elements of the workshop. One was good talks, both in the selection of topics and in length–in particular, the survey talks were highly appreciated. 78% of the respondents found the balance between longer and shorter talks mostly right, and 61% approved of the choice to have 55-minutes survey talks rather than 80-minutes tutorials. Another good aspect was the focused topic of the workshop, which made it easy to keep discussions relevant. Finally, the choice of participants was rated as balanced and conducive to a good atmosphere.

There was a general feeling, however, that the workshop program was perhaps a bit on the dense side, especially during the first one or two days.

When asked about topics that were felt to be missing, participants mostly cited neighbouring areas such as SAT solving, switching lemmas, and computational complexity theory in general, but some participants were also missing specific topics within proof complexity such as upper bounds for the Frege proof system and lower bounds for space complexity. It should be said, though, that the choice of topics for survey talks were based on an opinion poll before the workshop, and all topics with strong support in this opinion poll were given a survey talk slot (except when the organizing committee was unable to find a suitable speaker willing to give a survey talk).

As for the opposite question, whether some topics were superfluous, there was no clear consensus among the respondents, and the conclusion seems to be that for each topic a clear majority of participants felt that this topic was an essential one for the workshop. We had a combined panel discussion and open problems session, which 65% of the participants rated positively.

Regarding the social aspects of the seminar, participants were disappointed that there was not a hike, but felt it was a good decision to drop it because of bad weather. 89% of respondents enjoyed the music evening that was organized on Thursday.

To sum up, feedback was overwhelmingly positive. 83% of participants said they would definitely come again to a similar workshop, and 17% would probably come again.

## Participants


- Amirhossein Akbar Tabatabei
The Czech Academy of Sciences – Prague, CZ
- Albert Atserias
UPC – Barcelona, ES
- Paul Beame
University of Washington – Seattle, US
- Arnold Beckmann
Swansea University, GB
- Olaf Beyersdorff
University of Leeds, GB
- Ilario Bonacina
UPC – Barcelona, ES
- Igor Carboni Oliveira
University of Oxford, GB
- Marco Carmosino
University of California – San Diego, US
- Leroy Chew
University of Leeds, GB
- Stefan Dantchev
Durham University, GB
- Yuval Filmus
Technion – Haifa, IL
- Noah Fleming
University of Toronto, CA
- Michael A. Forbes
University of Illinois – Urbana-Champaign, US
- Nicola Galesi
Sapienza University of Rome, IT
- Michal Garlik
UPC – Barcelona, ES
- Joshua A. Grochow
University of Colorado – Boulder, US
- Tuomas Hakoniemi
UPC – Barcelona, ES
- Johan Hastad
KTH Royal Institute of Technology – Stockholm, SE
- Edward A. Hirsch
Steklov Institute – St. Petersburg, RU
- Pavel Hrubes
The Czech Academy of Sciences – Prague, CZ
- Dmitry Itsykson
Steklov Institute – St. Petersburg, RU
- Emil Jerabek
The Czech Academy of Sciences – Prague, CZ
- Jan Johannsen
LMU München, DE
- Raheleh Jalali Keshavarz
Czech Academy of Sciences – Brno, CZ
- Leszek Kolodziejczyk
University of Warsaw, PL
- Antonina Kolokolova
Memorial University of Newfoundland – St. John's, CA
- Oliver Kullmann
Swansea University, GB
- Massimo Lauria
Sapienza University of Rome, IT
- Meena Mahajan
Institute of Mathematical Sciences – Chennai, IN
- Barnaby Martin
Durham University, GB
- Moritz Müller
Universität Wien, AT
- Jakob Nordström
KTH Royal Institute of Technology – Stockholm, SE
- Joanna Ochremiak
University Paris-Diderot, FR
- Jan Pich
Universität Wien, AT
- Aaron Potechin
KTH Royal Institute of Technology – Stockholm, SE
- Pavel Pudlák
The Czech Academy of Sciences – Prague, CZ
- Ninad Rajgopal
University of Oxford, GB
- Kilian Risse
KTH Royal Institute of Technology – Stockholm, SE
- Robert Robere
University of Toronto, CA
- Rahul Santhanam
University of Oxford, GB
- Dmitry Sokolov
KTH Royal Institute of Technology – Stockholm, SE
- Neil Thapen
The Czech Academy of Sciences – Prague, CZ
- Jacobo Torán
Universität Ulm, DE
- Iddo Tzameret
Royal Holloway, University of London, GB
- Alasdair Urquhart
University of Toronto, CA
- Marc Vinyals
TIFR Mumbai, IN


Report from Dagstuhl Seminar 18052

# Genetic Improvement of Software

**Edited by**

**Justyna Petke[1], Claire Le Goues[2], Stephanie Forrest[3], and William B. Langdon[4]**

1 **University College London, GB,** `j.petke@ucl.ac.uk`
2 **Carnegie Mellon University, Pittsburgh, US,** `clegoues@cs.cmu.edu`
3 **Arizona State University, Tempe, US,** `stephanie.forrest@asu.edu`
4 **University College London, GB,** `w.langdon@cs.ucl.ac.uk`

## Abstract

We document the program and the immediate outcomes of Dagstuhl Seminar 18052 "Genetic Improvement of Software". The seminar brought together researchers in Genetic Improvement (GI) and related areas of software engineering to investigate what is achievable with current technology and the current impediments to progress and how GI can affect the software development process. Several talks covered the state-of-the-art and work in progress. Seven emergent topics have been identified ranging from the nature of the GI search space through benchmarking and practical applications. The seminar has already resulted in multiple research paper publications. Four by participants of the seminar will be presented at the GI workshop co-located with the top conference in software engineering - ICSE. Several researchers started new collaborations, results of which we hope to see in the near future.

**Seminar** January 28–February 2, 2018 – https://www.dagstuhl.de/18052
**2012 ACM Subject Classification** Software and its engineering → Automatic programming, Software and its engineering → Search-based software engineering
**Keywords and phrases** genetic improvement GI, search-based software engineering SBSE, software optimisation, evolutionary improvement, automated software improvement, automated program repair, evolutionary computation, genetic programming, GP
**Digital Object Identifier** 10.4230/DagRep.8.1.158
**Edited in cooperation with** Nicolas Harrand

## 1 Executive Summary

*Justyna Petke*
*Stephanie Forrest*
*William B. Langdon*
*Claire Le Goues*



Genetic improvement (GI) uses automated search to find improved versions of existing software. It can be used for improvement of both functional and non-functional properties of software. Much of the early success came from the field of automated program repair. However, GI has also been successfully used to optimise for efficiency, energy and memory consumption as well as automated transplantation of a piece of functionality from one program to another. These results are impressive especially given that genetic improvement only arose as a separate research area in the last few years. Thus the time was ripe to


Genetic Improvement of Software, *Dagstuhl Reports*, Vol. 8, Issue 01, pp. 158–182
Editors: Justyna Petke, Stephanie Forrest, William B. Langdon, and Claire Le Goues

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

organise a seminar that would gather researchers from GI and related areas together to summarise the current achievements and identify avenues for further research.

The seminar attracted researchers from various GI-related software engineering areas, ranging from automated software repair through genetic programming and software testing to biological and evolutionary computation. The talks covered the latest research and speculations on future research both in the practical applications of genetic improvement, such as energy consumption optimisation and automated parallelisation, to initial results on much lacking GI theory. In particular, GI theory and indeed software in general were discussed in terms of search landscape analysis. Other talks covered software testing and bug repair. The participants also identified a set of benchmarks and tools for GI. These have been published at the geneticimprovementofsofware.com website to allow other researchers to compare their new technologies against the state-of-the-art.

The seven breakout groups' topics ranged from re-evaluating the basic components of the GI framework, such as fitness functions and traversing the GI search space, to identifying issues related to adoption of GI in industry. One of the issues has been explanation of the automatically generated changes, which might be a roadblock in applying them in the real-world, especially safety-critical, software.

The seminar has already led to a few publications. For example, four papers accepted to the 4$^{\text{th}}$ International Genetic Improvement Workshop (GI-2018)[1], co-located with the International Conference on Software Engineering (ICSE), were written by one or more workshop participants. Indeed most were started in Dagstuhl. Several other collaborations have been established, with plans for visits and further research on topics identified at the seminar. We look forward to results of this work initiated at Dagstuhl.

## Introduction

Genetic improvement (GI) uses automated search to find improved versions of existing software [6, 8]. It uses optimisation, machine learning techniques, particularly search based software engineering techniques such as genetic programming [2, 1, 9]. to improve existing software. The improved program need not behave identically to the original. For example, automatic bug fixing improves program code by reducing or eliminating buggy behaviour, whilst automatic transplantation adds new functionality derived from elsewhere. In other cases the improved software should behave identically to the old version but is better because, for example: it runs faster, it uses less memory, it uses less energy or it runs on a different type of computer.

GI differs from, for example, formal program translation, in that it primarily verifies the behaviour of the new mutant version by running both the new and the old software on test inputs and comparing their output and performance in order to see if the new software can still do what is wanted of the original program and is now better. Using less constrained search allows not only functional improvements but also each search step is typically far cheaper, allowing GI to scale to substantial programs. Genetic improvement can be used to create large numbers of versions of programs, each tailored to be better for a particular use or for a particular computer, or indeed (e.g. to defeat the authors of computer viruses) simply to be different. Other cases where software need to be changed include porting to new environments (e.g. parallel computing [3] mobile devices) or for code obfuscation to prevent reverse engineering [7].

[1] http://geneticimprovementofsoftware.com/

Genetic improvement can by used with multi-objective optimisation to consider improving software along multiple dimensions or to consider trade-offs between several objectives, such as asking GI to evolve programs which trade speed against the quality of answers they give. Of course, it may be possible to find programs which are both faster and give better answers. Mostly Genetic Improvement makes typically small changes or edits (also known as mutations) to the program's source code, but sometimes the mutations are made to assembly code, byte code or binary machine code.

GI arose as a separate field of research only in the last few years. Even though it's origins could be traced back to the work by Ryan & Walsh [18] in 1995, it is the work by Arcuri [10] and White [20] that led to the development and wider uptake of the GI techniques. The novelty lay in applying heuristics to search for code mutations that improved existing software. Both Arcuri and White applied genetic programming (GP), with Arcuri using also hill-climbing and random search on a small set of problems. Rather than trying to evolve a program from scratch, as in traditional GP, Arcuri and White took the approach of seeding [5] the initial population with copies of the original program. Next, instead of focusing on evolving a program fulfilling a particular task, as has been done before, Arcuri and White used GP to improve their programs either to fix existing bugs or to improve the non-functional properties of software, in particular, its efficiency and energy consumption. Both Arcuri and White, however, applied their, now known as, GI techniques, to relatively small benchmarks having little resemblance to large scale real-world problems.

The bug fixing approach was taken up by Forrest, Le Goues and Weimer et al. [12, 15, 19] and adapted for large software systems. One of the insights that allowed for this adoption was an observation that full program variants need not be evolved, yet only a sequence of edits, which are then applied to the original program. Validity of the resultant modified software was then evaluated on a set of test cases, assumed to capture desired program behaviour, as in previous work. This strand of research led to the development of first GP-based automated software repair tool called GenProg [15]. Success of this automated bug fixing work led to several best paper awards and two 'Humie' awards (international prizes for human-competitive results produced by genetic and evolutionary computation http://www.human-competitive.org/) and inspired work on other automated software repair tools, including Angelix [16], which uses a form of constraint solving to synthesise bug fixes.

Research on improvement of non-functional software properties has yet to garner the attention and software development effort as the work on automated bug fixing. Langdon et al. [3, 13, 14] published several articles on efficiency improvement and parallelisation using GI. They were able to improve efficiency of large pieces of state-of-the-art software Moreover, the genetically improved version of a bioinformatics software called BarraCUDA is the first instance of a genetically improved piece of software adapted into development [14, 4].

Petke et al. [17] set themselves a challenge of improving efficiency of a highly-optimised piece of software that has been improved by expert human developers over a period of several years. In particular, a famous Boolean satisfiability (SAT) solver was chosen, called MiniSAT. It implements the core technologies of SAT solving and inspired a MiniSAT-hack track at the annual international SAT solver competitions, where anyone can submit their own version of MiniSAT. Petke et al. showed that further efficiency improvements can be made by using this source of genetic material for the GP process and specializing the solver for a particular downstream application. This work showed the initial potential of what is now called automated software transplantation and was awarded a Silver 'Humie'. Further work on automated software transplantation won an ACM SIGSOFT distinguished paper award and a Gold 'Humie' at this year's Genetic and Evolutionary Computation Conference (GECCO-2017) [11].

## Aims of the Seminar

The seminar brought together researchers in this new field of software engineering to investigate what is achievable with current technology and the current impediments to progress (if indeed there are any) of what can be achieved within the field in the future and how GI can affect the software development process.

With the growing popularity of the field, multiple awards and fast progress GI research in the field, it is the right time to gather top the academics in GI and related fields to push the boundaries of what genetic improvement can achieve even further.

This seminar brought researchers working in genetic improvement and related areas, such as automated program repair, software testing and genetic programming, together. It summarized achievements in automated software optimisation. We will use this summary as a basis to investigate how optimisation approaches from the different fields represented at the seminar can be combined to produce a robust industry-ready set of techniques for software improvement.

## References


1. Wolfgang Banzhaf, Peter Nordin, Robert E. Keller, and Frank D. Francone. *Genetic Programming – An Introduction; On the Automatic Evolution of Computer Programs and its Applications*. Morgan Kaufmann, San Francisco, CA, USA, January 1998.
2. John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA, 1992.
3. W. B. Langdon and M. Harman. Evolving a CUDA kernel from an nVidia template. In Pilar Sobrevilla, editor, *2010 IEEE World Congress on Computational Intelligence*, pages 2376–2383, Barcelona, 18–23 July 2010. IEEE.
4. W. B. Langdon and Brian Yee Hong Lam. Genetically improved BarraCUDA. *BioData Mining*, 20(28), 2 August 2017.
5. W. B. Langdon and J. P. Nordin. Seeding GP populations. In Riccardo Poli, Wolfgang Banzhaf, William B. Langdon, Julian F. Miller, Peter Nordin, and Terence C. Fogarty, editors, *Genetic Programming, Proceedings of EuroGP'2000*, volume 1802 of *LNCS*, pages 304–315, Edinburgh, 15–16 April 2000. Springer-Verlag.
6. William B. Langdon. Genetically improved software. In Amir H. Gandomi, Amir H. Alavi, and Conor Ryan, editors, *Handbook of Genetic Programming Applications*, chapter 8, pages 181–220. Springer, 2015.
7. Justyna Petke. Genetic improvement for code obfuscation. In Justyna Petke, David R. White, and Westley Weimer, editors, *Genetic Improvement 2016 Workshop*, pages 1135–1136, Denver, July 20–24 2016. ACM.
8. Justyna Petke, Saemundur O. Haraldsson, Mark Harman, William B. Langdon, David R. White, and John R. Woodward. Genetic improvement of software: a comprehensive survey. *IEEE Transactions on Evolutionary Computation*. In press.
9. Riccardo Poli, William B. Langdon, and Nicholas Freitag McPhee. *A field guide to genetic programming*. Published via `http://lulu.com` and freely available at `http://www.gp-field-guide.org.uk`, 2008. (With contributions by J. R. Koza).
10. Andrea Arcuri. Automatic software generation and improvement through search based techniques. PhD. Univ. of Birmingham, 2009.
11. Earl T. Barr, Mark Harman, Yue Jia, Alexandru Marginean, and Justyna Petke. Automated software transplantation. In *ISSTA*, pages 257–269, 2015.
12. Stephanie Forrest, ThanhVu Nguyen, Westley Weimer, and Claire Le Goues. A genetic programming approach to automated software repair. In *GECCO*, pages 947–954, 2009.
13. William B. Langdon and Mark Harman. Optimizing existing software with genetic programming. *IEEE Transactions on Evolutionary Computation*, 19(1):118–135, 2015.

14 William B. Langdon, Brian Yee Hong Lam, Justyna Petke, and Mark Harman. Improving CUDA DNA analysis software with genetic programming. In *GECCO*, pages 1063–1070, 2015.

15 Claire Le Goues, Michael Dewey-Vogt, Stephanie Forrest, and Westley Weimer. A systematic study of automated program repair: Fixing 55 out of 105 bugs for $8 each. In *ICSE*, pages 3–13, 2012.

16 Sergey Mechtaev, Jooyong Yi, and Abhik Roychoudhury. Angelix: scalable multiline program patch synthesis via symbolic analysis. In *ICSE*, pages 691–701, 2016.

17 Justyna Petke, Mark Harman, William B. Langdon, and Westley Weimer. Using genetic improvement and code transplants to specialise a C++ program to a problem class. In *EuroGP*, pages 137–149, 2014.

18 Paul Walsh and Conor Ryan. Automatic conversion of programs from serial to parallel using genetic programming - the Paragen system. In *ParCo*, pages 415–422, 1995.

19 Westley Weimer, ThanhVu Nguyen, Claire Le Goues, and Stephanie Forrest. Automatically finding patches using genetic programming. In *ICSE*, pages 364–374, 2009.

20 David R. White. Genetic programming for low-resource systems. PhD. Univ. of York, 2009.

## 2 Table of Contents

# 3 Overview of Talks

## 3.1 Progress in Structural Evolution for Bug Repair in JAVA

*Wolfgang Banzhaf (Michigan State University, US)*


**Joint work of** Wolfgang Banzhaf, Yuan Yuan (MSU CSE)
**Main reference** Yuan Yuan, Wolfgang Banzhaf: "ARJA: Automated Repair of Java Programs via Multi-Objective Genetic Programming", CoRR, Vol. abs/1712.07804, 2017.
**URL** http://arxiv.org/abs/1712.07804

Here we argue that (virtually) any structure can be made evolvable if one chooses the right elements of the structure and the proper rules of their combination, and provides sufficient guidance for the randomness of the mutation and crossover operators. We exemplify that argument by proposing a JAVA bug repair system that was inspired by GenProg, but further developed and adapted to JAVA. Results show the efficacy of evolutionary search over random search, multi-objective optimisation over single objective optimisation and "knowledge-enhanced" (smart) operators over others. A new set of bugs from the Defects4J benchmark suit can be successfully repaired, including multi-location bugs.

## 3.2 Automatic Parallelisation of Software Using Genetic Improvement

*Bobby R. Bruce (University College London, GB)*


**Joint work of** Bobby R. Bruce, Justyna Petke

While the use of hardware accelerators, like GPUs, can significantly improve software performance, developers often lack the expertise or time to properly translate source code to do so. We highlight two approaches to automatically offload computationally intensive tasks to a system's GPU by generating and inserting OpenACC directives; one using grammar-based genetic programming, and another using a bespoke four stage process. We find that the grammar-based genetic programming approach is capable of reducing execution time by 2.60% on average, across the applications studied, while the bespoke four-stage approach reduces execution time by 2.44%.

However, our investigation shows a handwritten OpenACC implementation is capable of reducing execution time by 65.68%, suggesting our techniques could be improved upon. Comparing the differences, we find our techniques do not handle data to and from the GPU in a sensible manner and that, if they did, considerably execution time savings are possible. We therefore advise future researchers to focus on the automation of transferring data between main and GPU memory; a problem search-based software engineering is capable of solving.

## 3.3 Assuring Organic Programs

*Myra B. Cohen (University of Nebraska, Lincoln, US)*


**Joint work of** Myra, B. Cohen, Justin Firestone, Massimiliano Pierobon
**Main reference** Myra B. Cohen, Justin Firestone, Massimiliano Pierobon: "The Assurance Timeline: Building Assurance Cases for Synthetic Biology", in Proc. of the Computer Safety, Reliability, and Security - SAFECOMP 2016 Workshops, ASSURE, DECSoS, SASSUR, and TIPS, Trondheim, Norway, September 20, 2016, Proceedings, Lecture Notes in Computer Science, Vol. 9923, pp. 75–86, Springer, 2016.
**URL** https://doi.org/10.1007/978-3-319-45480-1_7

Recent research in genetic improvement and self-adaptation have created a class of programs that we call organic, since they follow an evolution cycle similar to that of living organisms. Traditional testing techniques assume that program modifications are planned, systematic and well understood. However, this may not be true for organic programs. I discuss the use of an assurance case to argue about the dependability and safety of an organic program using an exemplar from synthetic biology (which are living organic programs). I present an orthogonal dimension to an assurance case, the assurance timeline, which aims to reason about the dynamic, evolving aspects of these systems.

## 3.4 Genetic-Improvement of Test suite

*Benjamin Danglot (INRIA Lille, FR)*



In the literature there is a rather clear segregation between tests manually written by developers and automatically generated ones. DSpot explores a third solution: automatically improving existing test cases written by developers. DSpot takes as input developer-written tests and synthesizes an improved version. Those improvements are given to the developer as a pull-request than can be directly integrated into their code-base. DSpot uses mutation operators on the code of each test, it produces assertions and selects them according to a given test criterion such as coverage. In 26/40 cases, DSpot has been able to create a better version of a test class. We proposed pull requests to real developers and 7 of them have been added permanently to their test suite.

## 3.5 Software Plasticity

*Nicolas Harrand (KTH Royal Institute of Technology, Stockholm, SE)*


**Joint work of** Benoit Baudry, Nicolas Harrand

Approximate computing, automatic diversification and genetic improvement are techniques that all rely on *speculative transformations*: transformations that aim at producing variants of a program that are functionally similar to the original, yet execute slightly differently. The intuition of all the techniques cited above potential enhancements lie in these acceptable behavioural differences (enhanced performance, security, reliability, etc.).

The design of the speculative transformations that can yield these improvements remains a critical challenge. These transformations must target regions of programs that can tolerate changes in the execution flow, while maintaining the correctness of the program. We call them *plastic code regions*. We contribute with fundamental new knowledge about these regions in object-oriented programs, as well as with new kinds of speculative transformations that directly exploit this new knowledge.

Our empirical inquiry of plastic code regions starts from a random exploration of three classical speculative transformations: add, replace and delete statements. We synthesize 24 583 variants from 6 real-world Java programs, and focus our analysis on the 5305 that are similar, modulo test suite, to the original. Our key insights about plastic regions are as follows: developers naturally write code that supports fine-grain behavioural changes; statement deletion is a surprisingly effective; high-level design decisions, such as the choice of a data structure, are natural points that can evolve while keeping functionality. Based on these new findings, we design targeted speculative transformations and show that they are very effective at producing variants that are both similar (modulo tests) and different from the original.

## 3.6 DeepBugs: Learning to Find Bugs

*Michael Pradel (TU Darmstadt, DE)*


**Joint work of** Michael Pradel, Koushik Sen

Automated bug detection, e.g., through pattern-based static analysis, is an increasingly popular technique to find programming errors and other code quality issues. Traditionally, bug detectors are program analyses that are manually written and carefully tuned by an analysis expert. Unfortunately, the huge amount of possible bug patterns makes it difficult to cover more than a small fraction of all bugs. I present a new approach toward creating bug detectors. The basic idea is to replace manually writing a program analysis with training a machine learning model that distinguishes buggy from non-buggy code. To address the challenge that effective learning requires both positive and negative training examples, we use simple code transformations that create likely incorrect code from existing code examples. We present a general framework, called DeepBugs, that extracts positive training examples from a code corpus, leverages simple program transformations to create negative training examples, trains a model to distinguish these two, and then uses the trained model for identifying programming mistakes in previously unseen code. As a proof of concept, we create four bug detectors for JavaScript that find a diverse set of programming mistakes, e.g., accidentally swapped function arguments, incorrect assignments, and incorrect binary operations. To find bugs, the trained models use information that is usually discarded by program analyses, such as identifier names of variables and functions. Applying the approach to a corpus of 150,000 JavaScript files shows that learned bug detectors have a high accuracy, are very efficient, and reveal 132 programming mistakes in real-world code.

## 3.7 Analyzing Neutrality in Program Space

*Joseph Renzullo (Arizona State University, Tempe, US)*


**URL** https://docs.google.com/presentation/d/18-0b4Mdnvum28IRCCLUb966Gb2VDHUf0D88DhlmakGI/edit?usp=sharing

I present evidence of interaction between multiple edits (both positive and negative epistasis) in the region near the original program. There are a few cases where repairs were found which were attributed to multiple independent patches working in combination (previous results have shown that these often minimise to one patch) here we show evidence that this is not always the case.

Additionally, I raise questions about how methods in biology (particularly borrowing from theoretical biology) may be used to characterise (and hopefully exploit) the topology of neutral space.

## 3.8 Approximate computing

*Lukas Sekanina (Brno University of Technology, CZ)*


**Joint work of** Lukas Sekanina, Zdenek Vasicek, Vojtech Mrazek
**Main reference** Vojtech Mrazek, Syed Shakib Sarwar, Lukás Sekanina, Zdenek Vasícek, Kaushik Roy: "Design of power-efficient approximate multipliers for approximate artificial neural networks", in Proc. of the 35th International Conference on Computer-Aided Design, ICCAD 2016, Austin, TX, USA, November 7-10, 2016, p. 81, ACM, 2016.
**URL** http://dx.doi.org/10.1145/2966986.2967021

A new design paradigm–approximate computing–was established to investigate how computer systems can be made better (e.g. more energy efficient, faster, and less complex) by relaxing the requirement that they are exactly correct. We provide a brief introduction to approximate computing and indicates how evolutionary computation, in general, and genetic improvement, in particular, can be employed to provide requested approximations. An important case study is presented in the area of evolutionary approximation of multipliers (which are key to performance) for deep neural networks.

## 3.9 An Actionable Performance Profiling for JavaScript

*Marija Selakovic (TU Darmstadt, DE)*



Many programs suffer from performance problems, but unfortunately, finding and fixing such problems is a cumbersome and time-consuming process. My work focuses on JavaScript, for which little is known about performance issues and how developers address them. To address these questions, I present the main findings from the empirical study of $\approx 100$ reproduced performance-related issues from popular JavaScript projects. To help developers find and fix recurrent performance issues I present two profiling approaches. The first approach focuses on detecting finding the optimal order of checks in logical expressions and switch statements

and proposing beneficial changes to the developers. Optimizing the order of evaluations reduces the execution time of individual functions by between 2.5% and 59%, and leads to statistically significant application-level performance improvements that range between 2.5% and 6.5%. The second approach helps developers find and fix performance problems related to API usages. The technique focuses on finding conditionally-equivalent but performance-wise different APIs. Our evaluation with 939 APIs from 8 popular JavaScript libraries shows the prevalence of conditionally equivalent APIs. In particular, out of 217 API pairs that are equivalent for a subset of all inputs, our technique derives an equivalence condition for 149 pairs. Furthermore, it finds that 147 API pairs have different performance, enabling developers to exploit conditional equivalences to speed up their code.

## 3.10 Repairing crashes in Android apps

*Shin Hwei Tan (National University of Singapore, SG)*


**Joint work of** Shin Hwei Tan, Zhen Dong, Xiang Gao, Abhik Roychoudhury
**Main reference** Shin Hwei Tan, Zhen Dong, Xiang Gao, Abhik Roychoudhury: "Repairing Crashes in Android Apps", in Proc. of the ACM/IEEE Int'l Conf. on Software Engineering (ICSE).,To Appear, 2018.

Android apps are omnipresent, and frequently suffer from crashes. This leads to poor user experience and loss of revenue. Past work has focused on automated test generation to detect crashes in Android apps. However automated repair of crashes has not been studied. We propose the first approach to automatically repair Android apps, specifically we propose a technique for fixing crashes in Android apps. Unlike most test-based repair approaches, we do not need a test-suite; instead a single failing test is meticulously analyzed for crash locations and reasons behind these crashes.Unlike most test-based repair approaches, we do not need a test-suite; instead a single failing test is meticulously analyzed for crash locations and reasons behind these crashes. Our approach hinges on a careful empirical study which seeks to establish common root-causes for crashes in Android apps, and then distils the remedy of these root-causes in the form of eight generic transformation operators. These operators are applied using a search-based repair framework embodied in our repair tool *Droix*. We also prepare a benchmark *DroixBench* capturing reproducible crashes in Android apps. Our evaluation of *Droix* on DroixBench reveals that the automatically produced patches are often syntactically identical to the human patch, and on some rare occasion even better than the human patch (in terms of avoiding regressions). These results confirm our intuition that our proposed transformations form a sufficient set of operators to patch crashes in Android.

## 3.11 BugZoo: A platform for studying historical bugs

*Christopher Timperley (Carnegie Mellon University, Pittsburgh, US)*


**URL** https://github.com/squaresLab/BugZoo

I introduce BugZoo to the genetic improvement community: BugZoo is an open-source platform for studying historical software bugs that helps researchers to conduct high-quality reproducible experiments. BugZoo can be used to conduct experiments in a diversity of

fields including but not limited to software testing, program repair, genetic improvement, fault localisation, and program analysis. By providing a rich API, a decentralised means of distribution bugs, and a controlled execution environment, BugZoo makes it faster and easier to perform research.

## 3.12 Major Transitions in Information Technology

*Sergi Valverde (UPF, Barcelona, ES)*


**Main reference** Sergi Valverde: "Major Transitions in Information Technology", Phil. Trans. R. Soc. B, 371: 20150450, 2016.
**URL** http://dx.doi.org/10.1098/rstb.2015.0450

When looking at the history of technology, we can see that all inventions are not of equal importance. Only a few technologies have the potential to start a new branching series (specifically, by increasing diversity), have a lasting impact in human life and ultimately became turning points. Technological transitions correspond to times and places in the past when a large number of novel artefact forms or behaviours appeared together or in rapid succession. Why does that happen? Is technological change continuous and gradual or does it occur in sudden leaps and bounds? The evolution of information technology (IT) allows for a quantitative and theoretical approach to technological transitions. The value of information systems experiences sudden changes (i) when we learn how to use this technology, (ii) when we accumulate a large amount of information, and (iii) when communities of practice create and exchange free information. The coexistence between gradual improvements and discontinuous technological change is a consequence of the asymmetric relationship between complexity and hardware and software. Using a cultural evolution approach, we suggest that sudden changes in the organization of ITs depend on the high costs of maintaining and transmitting reliable information.

# 4 Working groups

## 4.1 Pseudo Neutrality

*Benoit Baudry (KTH Royal Institute of Technology, Stockholm, SE)*



The synthesis of pseudo-neutral program variants is a key challenge for genetic improvement. Given an original program that one wishes to improve, pseudo-neutral variants are those programs that are functionally similar to the original, yet exhibit some differences in their behaviour. More precisely, given a program P that passes all the tests in $TS$, we wish to generate variants of $P$ that are synthesised by transformation $\tau$ and that are such that

- $\tau(P)$ passes all tests in $TS$, i.e., $P$ and $\tau(P)$ are equivalent modulo $TS$
- $\{traces\}_P \neq \{traces\}_{\tau(P)}$ i.e., $P$ and $\tau(P)$ are semantically different

This definition is summarised in Figure 1
Key insight: pseudo-neutral program variants exist!

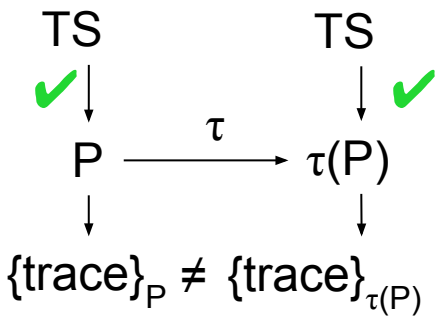



**Figure 1** Pseudo-neutral program variants.

We have strong empirical evidence of their existence in large quantities and in many languages (Java, C and assembly code) [1, 2]. Our results also demonstrate that the existence of these variants is independent of the strength of the test suite $TS$ that used to assess the functional similarity between variants.

It is important to note that these program variants are not equivalent mutants in the sense of mutation testing. Indeed, as illustrated in Figure 1, the execution traces vary between the original and the transformed program. This means that the behaviour are not equivalent and that the transformed program is not equivalent to the original.

### 4.1.1 Challenges

The variants and the original must have different traces for the same input. Yet, there are many ways to capture the traces (function calls, system calls, states, etc.). The question about what is the most appropriate or relevant level to capture traces is still open.

The synthesis of variants relies on transformations on the code (syntactic changes), yet the goal is to produce semantic variations. One challenge here is to know how to predict the semantic impact of a syntactic change. A similar questions is to know what makes software prone to the existence of these pseudo mutational robust variants.

Can we design an experiment to explore whether the following biological phenomenon holds in software: adding levels of complexity enhances robustness and evolvability in a multilevel genotype-phenotype map.

### References

1 Benoit Baudry, Simon Allier, and Martin Monperrus. Tailored source code transformations to synthesize computationally diverse program variants. In *Proc. of the Int. Symp. on Software Testing and Analysis (ISSTA)*, pages 149–159, CA, USA, 2014.

2 Eric Schulte, Zachary Fry, Ethan Fast, Westley Weimer, and Stephanie Forrest. Software mutational robustness. *Genetic Programming and Evolvable Machines*, 15(3):281–312, September 2014.

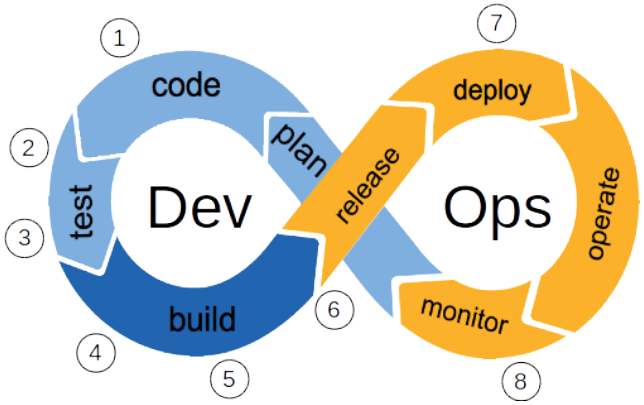



**Figure 2** DevOps Software life cycle.

## 4.2 Genetic Improvement for DevOps

*Nicolas Harrand (KTH Royal Institute of Technology, Stockholm, SE)*



We discussed the opportunities to deploy genetic improvement offered by DevOps [2]. DevOps aims to close the loop between development and operation of software. To achieve this goal, it relies heavily on automation of the software construction process. In this inclination toward automation and the resulting problems, lie many opportunity to integrate Genetic Improvement into the software construction pipeline. Among them, we identified the following:

1. Fixing merge conflicts
2. Genetically improve the test suite
3. Use GI to fix flaky tests
4. Integrate bug repair into Continuous Integration (CI) tools.
5. Automatic fixes in dependency conflicts
6. Container minimization
7. Deployment of a diverse population of software
8. Use of monitoring feedback for further improvements

The discussion resulted in the publication of a workshop paper listing and detailing these different opportunities [1].

### References

1 Benoit Baudry, Nicolas Harrand, Eric Schulte, Chris Timperley, Shin Hwei Tan, Marija Selakovic, Emamurho Ugherughe A spoonful of DevOps helps the GI go down. In *Proceedings of Workshop on Genetic Improvement (GI 2018)*, pp. 35–37, Gothenburg, Sweden, June 2018. ACM.

2 Christof Ebert, Gorka Gallardo, Josune Hernantes, and Nicolas Serrano. Devops. *IEEE Software*, 33(3):94–100, 2016.

## 4.3 Benchmarks and Corpora

*Myra B. Cohen (University of Nebraska, Lincoln, US), William B. Langdon (University College London, GB), and Claire Le Goues (Carnegie Mellon University, Pittsburgh, US)*



This working group discussed the need for benchmarks and a corpus of programs that have been created through genetic improvement. The group came up with two dimensions of this problem, (1) **benchmarking** and (2) **building a corpus**:

1. **Benchmarking**, i.e. providing programs and example bugs/functionality, etc. for others to evaluate their GI techniques. Some benchmarks already exist.
   Potential issue: There is a risk that people will overfit their techniques to these benchmarks.
   Types of Artefacts that we should collect:
   - Failing and passing test cases (or other witnesses of desired/undesired behaviour)
   - Program with bugs and set of properties of interest
   - Patches: This would include a patch and undo approach (always return to base model to re-patch)
2. **Corpus**, i.e. providing the artefacts from GI throughout a program's history which includes the program, the patches, the new programs, etc.
   Some people may not want to re-run the programs and build their own artefacts. This provides programs and other GI artefacts for them to study.
   **Artefacts:**
   - Base program
   - History of the evolved program and patches
   - This assumes each patch builds on another
   - Includes patches for bugs, optimisation and transplantation
   - Can be used to mine information, understand the artefacts

## 4.4 Energy Breakout Session

*Markus Wagner (University of Adelaide, AU)*



The minimisation of energy consumption is an important challenge in many domains. We focussed on two domains: electric circuits and mobile phones.

Researchers who develop new circuits can often rely on good models. Some are lightweight and provide estimates based on switching activity and gate size analysis, and they are often good enough to reliably drive tournament selection. For the final functional validation, either an actual implementation is used, or SAT solvers. Vasicek and Sekanina [1] applied simple and so cheap area and delay estimation techniques during the evolutionary approximation of digital circuits. Parameters of best-evolved circuits were then verified by means of a professional circuit design tool. The quality of estimated values was sufficient for their purposes. Mrazek et al. [2] applied these estimation techniques in evolutionary design/improvement of specialised multipliers for deep neural networks.

When dealing with mobile phones, there are software engineering issues to solve before getting to the optimisation problem. In addition, complex interactions on the actual phone make it difficult to consistently see the benefit of a change. Possible optimisation approaches range from working on the hardware (voltage schedules, frequency adjustments) to code changes. The group discussed various targets: screen, communication, GPS, and code. It was not clear to the group if hardware-in-the-loop is necessary for the evolution, although the group identified cases were the creation of sufficiently precise models is out of question. When it comes to in-vivo optimisation, then the processes need to deal with large amounts of noise from various resources. This is immensely prevalent when attempting to optimise communication. Also, the use of external power meters is becoming increasingly difficult as the phone's communication with the battery is hard to mimic. Bokhari et al. [3,4] characterised noise and challenges, and performed multi-objective configuration optimisation on Android 6 devices.

In-vivo optimisation is interesting when the target device's exact configuration is now known. Recently, Yoo et al. [5,6] demonstrated that this is possible for performance optimisation.

### References

1 Vasicek Zdenek and Sekanina Lukas. Circuit Approximation Using Single and Multi-Objective Cartesian GP. In: EuroGP. Springer, 2015, pp. 217–229.
2 Mrazek Vojtech, Sarwar Syed Shakib, Sekanina Lukas, Vasicek Zdenek and Roy Kaushik. Design of Power-Efficient Approximate Multipliers for Approximate Artificial Neural Networks. In: Proceedings of the IEEE/ACM International Conference on Computer-Aided Design. Austin, USA, 2016, pp. 811–817.
3 Mahmoud A. Bokhari, Bobby R. Bruce, Brad Alexander, and Markus Wagner. 2017. Deep parameter optimisation on Android smartphones for energy minimisation: a tale of woe and a proof-of-concept. In GECCO-2017. pp. 1501–1508.
4 Mahmoud A. Bokhari, Yuanzhong Xia, Bo Zhou, Brad Alexander, Markus Wagner. Validation of Internal Meters of Mobile Android Devices. 2017. https://arxiv.org/abs/1701.07095
5 Jeongju Sohn and Seongmin Lee and Shin Yoo. Deep Parameter Optimisation of GPGPU Work Group Size for OpenCV. In: SSBSE 2016, LNCS 9962, 211–217. Springer.
6 Shin Yoo. Amortised Optimisation of Non-functional Properties in Production Environments. In: SSBSE 2015, LNCS 9275, pp 31–46. Springer.

## 4.5 Diversity

*Wolfgang Banzhaf (Michigan State University, US)*



With the prevalence of neutrality in computer code, we agreed that the more important issue is how to create diversity in a population. As we start evolution by a working program that needs to be improved, this is an issues, since we come from a situation where there is a solution, but one which we want to further improve on. So, how do we create diversity, since we are not allowed to just randomly create programs?

There was discussion about the fact that the neutral networks are actually quite intricately connected. So would diversity actually be so important?

As far as neutrality is concerned, many mentioned the issue of really very flat fitness landscapes. Where would there be a signal for improvements? Two measures where emphasised for avoiding getting stuck on the plain:

1. Random subset selection
2. Co-evolutionary approaches

## 4.6 Comprehensibility and Explanation

*Colin Johnson (University of Kent, UK)*



An important issue in applying genetic improvement in practical software development is convincing developers to take up the improvements suggested by GI systems. This can be tackled in a number of different ways. For example, running the modified programs on test data can be used both to check whether test cases are still satisfied post-improvement, and to measure improvements to non-functional properties. Another approach is to apply static analysis and verification techniques to GI-modified programs to examine properties. A third approach, which we focus on here, is that of making modified code comprehensible to human programmers, and for the GI system to provide human-comprehensible explanations and annotations for developers.

### 4.6.1 Human Readability

One way to make improvements convincing is to make changes so that a human programmer can easily read the results from the suggested improvement. This could in part be achieved by keeping code changes small and focused (perhaps only altering code regions specified by the developer), and avoiding side-effects of genetic operations such as code bloat. A related, almost opposite, issue, is avoiding the GI system making excessively "clever" convoluted improvements that might use unusual language or API features or use language in a non-idiomatic way. One approach to this would use some notion of robustness, i.e. measuring whether syntactically-similar programs have similar behaviour. Another might use an approach inspired by economics or ecology, giving the GI system a fixed budget of changes to use. A final approach might be capturing, measuring, and optimising for the notion of idiomatic code, the kind of code that humans use.

### 4.6.2 Explainability

Another approach to making improvements convincing is for the GI system to generate an explanation for the improvement alongside the improvement itself. At a simple level, this explanation could consist of giving some examples that exemplify the improvement; for example, in a fault-fixing system, examples of test cases that are now satisfied that weren't before the fix. A harder challenge is to provide a higher-level explanation of the improvements made, particularly having the system explain the overall effect of many small changes to the code. This might come from some analysis of the improvement process, or by some post-improvement comparative analysis of the improved code against the original code.

### 4.6.3 Comprehensibility as Improvement

Alternatively, human-comprehensibility might be the aim of a GI process. A GI process might aim to refactor code to bring it into a common style, for example in the use of consistent names, common code idioms, exception handling. Another related area, which had already been explored somewhat in the literature, is optimising code against measures of code complexity, for example cohesion and coupling in Object-Oriented systems. A related idea would be to use GI to refactor code to use common design patterns. Another issue is about changing the granularity of code: breaking a single expression into sub-components to allow more fine grained change/tuning; or, abstracting away from detailed code into a higher-level framework, replacing detailed code with a macro or API call.

### 4.6.4 Trade-offs

We can consider how we might trade off comprehensibility against other properties, particularly non-functional properties. Perhaps, given a sufficiently reliable GI system, we could see a system that allowed rapid automated refactoring of code: for example, a human-readable piece of code being transformed into an energy-efficient one for deployment, then changed back into a human-comprehensible one for a developer to make improvements, then into a more evolvable structure for the computer to make different kinds of improvements, ... Finally, there is the difficult issue of the effect of GI and other code-transformation methods on developer's mental models of code. However readable the code is in isolation, there is still the issue of how much a set of changes breaks a specific developer's understanding of how their specific piece of code works.

#### References


1 Raymond P.L. Buse and Westley R. Weimer. Automatically documenting program changes. In *Proceedings of the IEEE/ACM International Conference on Automated Software Engineering*, ASE '10, pages 33–42, New York, NY, USA, 2010. ACM.

2 Abram Hindle, Earl T. Barr, Mark Gabel, Zhendong Su, and Premkumar Devanbu. On the naturalness of software. *Commun. ACM*, 59(5):122–131, April 2016.

3 Abram Hindle, Earl T. Barr, Zhendong Su, Mark Gabel, and Premkumar Devanbu. On the naturalness of software. In *Proceedings of the 34th International Conference on Software Engineering*, ICSE '12, pages 837–847, Piscataway, NJ, USA, 2012. IEEE Press.

4 Katsuhisa Maruyama and Ken-ichi Shima. Automatic method refactoring using weighted dependence graphs. In *Proceedings of the 21st International Conference on Software Engineering*, ICSE '99, pages 236–245, New York, NY, USA, 1999. ACM.

5 Mark O'Keeffe and Mel O Cinneide. Search-based software maintenance. In *Proceedings of the Conference on Software Maintenance and Reengineering*, CSMR '06, pages 249–260, Washington, DC, USA, 2006. IEEE Computer Society.

6 Ali Ouni, Marouane Kessentini, Mel O Cinneide, Houari Sahraoui, Kalyanmoy Deb, and Katsuro Inoue. More: A multi-objective refactoring recommendation approach to introducing design patterns and fixing code smells. *Journal of Software: Evolution and Process*, 29(5):e1843.

7 Baishakhi Ray, Vincent Hellendoorn, Saheel Godhane, Zhaopeng Tu, Alberto Bacchelli, and Premkumar Devanbu. On the "naturalness" of buggy code. In *Proceedings of the 38th International Conference on Software Engineering*, ICSE '16, pages 428–439, New York, NY, USA, 2016. ACM.

8 Tushar Sharma. Identifying extract-method refactoring candidates automatically. In *Proceedings of the Fifth Workshop on Refactoring Tools*, WRT '12, pages 50–53, New York, NY, USA, 2012. ACM.

## 4.7 Fitness Functions for Genetic Improvement

*Brad Alexander (University of Adelaide, AU)*



### 4.7.1 Test suite selection

This topic examines the potential to make use of improved test suites and other data to provide for faster and better-informed search. The motivations of this topic are observed problems in terms flat fitness landscapes for some applications such as defect repair. More generally, there is also the problem that full evaluation all tests for each fitness evaluation impacts on the speed of search.

#### 4.7.1.1 Reducing evaluation times

One approach (Langdon) to improving the speed of each evaluation is to select a representative subset of tests for each fitness evaluation [1]. For the GI objective of reducing execution time the representative subset might be three tests, one short-lived, one of intermediate length, and one long-running. Run the short lived one first and, if that fails, don't necessarily bother with the others.

Questions arising from this approach include by how much this improves the speed of search? Some of the trade-offs have already been studied in: [3]. This showed that, when coupled with a more informed fitness landscape, being selective in the tests run can greatly speed search in the domain of defect repair (more on this below). Other suggestions included using a steady state GA to minimise the number of fitness evaluations [8].

More broadly, some proposals for choosing representative test samples included (Joseph) Eigentest weighting for the most representative sample of tests. (Celso) The literature on test selection in search-based software engineering (SBSE) is quite strong [6]. There is still work to be done on productive strategies to use for particular GI objectives.

Another approach to minimizing test suite evaluations time is to only apply tests that exercise the code that is changed by a patch. This approach is used in industry. Is it used in GI?

#### 4.7.1.2 Boosting approaches

For the objective of defect repair (Joseph) one approach is to favor subsets of tests that are most likely to fail at the current stage of search. This helps shape the search toward overcoming challenging cases first. This boosting approach has been long-used in Genetic Programming [7]. This approach has also been applied in the improvement of search spaces in automated program repair [9].

There was also some discussion of the potential benefits of applying subsets of tests to individuals in terms of preserving useful genetic material. That is, individuals that might, when applied to all tests, achieve low fitness could still have useful materials that contributes to solutions through its progeny. It has been observed (Joseph) that such individuals, if allowed to persist can act as repositories for useful material – if the sub-set of tests two which they are exposed allow them to survive. This approach mirrors Lee Spector's Lexicase testing approach – randomly select a test case to select parents – only if there is a tie do we select a second case and so on (see: [5] for recent study).

#### 4.7.1.3 Test feature selection

Questions arising from this approach include by how much this improves the speed of search? Is this approach sensitive to the selection of test features. In this context, a test feature is some intrinsic or manifest property of the test. Examples of test features might be: the length of time that it takes to run a test; the current likelihood that a given test will fail relative to the population of variants; and the coverage spectrum of a particular test.

#### 4.7.1.4 Specialised domains

Improving tests involving GUI interactions, (e.g. for mobile devices) is a concern for GI objectives such as energy optimisation. The Monkey test generator generates shallow traversals of GUI interfaces due it its unguided nature. In contrast Sapienz does well in traversing through interfaces of mobile devices but can sometimes aggressively generate states that app programmers might consider infeasible. GUI ripping (e.g. [2]) can provide some help in this regard.

### 4.7.2 Landscape Improvement

For some GI objectives such as defect repair the landscape can be very flat and uninformative. Some work such as [3] automatically derived program invariants and was able to leverage these to speed up search. More recent work in program invariants [2] has been used for fault localisation – perhaps this can be leveraged both for better localisation of repair locations but also for giving a more informed fitness response to programs. Evosuite [4] continues to be developed as a way to reverse engineer assertions that serve as oracles from which to generate tests (the approach is contingent on the assumption that the version of the program used to generate tests is correct). The extent to which tests based on invariants can be synthesised from programs with bugs is an interesting question.

Another approach is the use of smart operators that are more likely to preserve semantics. Elements of this approach, combined with the use of genotypes that allow for separate evolution of the source, destination, and operation in a defect repair setting have been recently applied with some success in AJAR[10].

### References

1 William B. Langdon and Mark Harman. Optimising existing software with genetic programming. *IEEE Transactions on Evolutionary Computation*, 19(1):118–135, February 2015. doi:10.1109/TEVC.2013.2281544.

2 Domenico Amalfitano, Anna Rita Fasolino, Porfirio Tramontana, Salvatore De Carmine, and Atif M. Memon. Using gui ripping for automated testing of android applications. In *Proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering*, ASE 2012, pages 258–261, New York, NY, USA, 2012. ACM.

3 Ethan Fast, Claire Le Goues, Stephanie Forrest, and Westley Weimer. Designing better fitness functions for automated program repair. In *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*, GECCO '10, pages 965–972, 2010. ACM.

4 Gordon Fraser and Andrea Arcuri. Evosuite: Automatic test suite generation for object-oriented software. In *Proceedings of the 19th ACM SIGSOFT Symposium and the 13th European Conference on Foundations of Software Engineering*, ESEC/FSE '11, pages 416–419, New York, NY, USA, 2011. ACM.

5 Ting Hu and Karoliina Oksanen. Lexicase selection promotes effective search and behavioural diversity of solutions in linear genetic programming. CEC-2017

6 Yuanyuan Zhang Mark Harman, Yue Jia. Achievements, open problems and challenges for search based software testing.
7 Gregory Paris, Denis Robilliard, and Cyril Fonlupt. Applying boosting techniques to genetic programming. In Pierre Collet, Cyril Fonlupt, Jin-Kao Hao, Evelyne Lutton, and Marc Schoenauer, editors, *Artificial Evolution*, pages 267–278, 2002. Springer Berlin Heidelberg.
8 Eric Schulte, Jonathan Dorn, Stephen Harding, Stephanie Forrest, and Westley Weimer. Post-compiler software optimization for reducing energy. *SIGARCH Comput. Archit. News*, 42(1):639–652, February 2014.
9 Stephanie Forrest Westley Weimer, Zachary P. Fry. Leveraging program equivalence for adaptive program repair: Models and first results. In *2013 28th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pages 356–366, 2013. doi:10.1109/ASE.2013.6693094.
10 Yuan Yuan and Wolfgang Banzhaf. ARJA: Automated repair of java programs via multi-objective genetic programming. *arXiv:1712.07804*.

## 5 Resources for Genetic Improvement

The seminar was also the occasion for participants to share useful resources such as state-of-the-art tools and frameworks etc. In particular, since the evaluation of new techniques in a replicable and time efficient fashion may prove challenging, the exchange of benchmarks is a valuable output of the seminar.

### 5.1 Tools, Libraries and Frameworks

1. DSpot: a tool for Genetic Improvement of test suites
   https://github.com/STAMP-project/dspot
2. PyGGI: Python General Framework for GI
   https://github.com/coinse/pyggi
   https://coinse.github.io/pyggi/ (API documentation)
3. GIN: GI in no Time
   https://github.com/gintool/gin
4. GenProg:
   https://squareslab.github.io/genprog-code/ GitHub io page,
   https://github.com/squaresLab/genprog-code GitHub source
5. Software Engineering Library (support for C/C++ source w/ CLANG, ASM, ELF, future: Coq, Java):
   https://github.com/GrammaTech/sel Github Source
   https://grammatech.github.io/sel/Manual
   https://grammatech.github.io/sel/Usage.html Installation and easy examples to start
   https://github.com/GrammaTech/clang-mutate C/C++ manipulation tooling
6. ARJA:
   https://github.com/yyxhdy/arja
7. Reproduce and repair failing builds:
   https://github.com/Spirals-Team/librepair/tree/master/repairnator
8. MuScalpel: automated software transplantation.
   http://crest.cs.ucl.ac.uk/autotransplantation/downloads/muScalpel.zip

9. History of programming languages
   https://github.com/svalver/Proglang
10. Agent-based model for the cultural diffusion of programming languages (code)
    http://modelingcommons.org/browse/one_model/4611
11. JavaScript parser:
    http://esprima.org/
    https://github.com/estools/escodegen
    https://github.com/estools/estraverse
12. Astor4Android: program repair for Android App
    https://github.com/kayquesousa/astor4android

## 5.2 Benchmarks

1. BugZoo (Docker containers for ManyBugs): https://github.com/squaresLab/BugZoo
2. CodeFlaws
   https://github.com/codeflaws/codeflaws
3. Parsec:
   http://parsec.cs.princeton.edu/
4. SPEC INT:
   https://www.spec.org/benchmarks.html
5. Microsoft Version Control repos with Bug Info related to commits:
   http://msr.uwaterloo.ca/msr2009/challenge/msrchallengedata.html
6. DBGBENCH : evaluation of automated fault localization, diagnosis, and repair techniques w.r.t. the judgement of human experts
   https://github.com/rjust/defects4j a collection of reproducible bugs
   https://droix2017.github.io a set of reproducible crashes in Android apps
7. ARJA Benchmark of seed bugs
   https://github.com/yyxhdy/SeededBugs

# 6 Following the Seminar, New work and New Connections

## 6.1 New Work

### References


1 Afsoon Afzal, Jeremy Lacomis, Claire Le Goues, and Christopher S. Timperley. A Turing test for genetic improvement. In Justyna Petke, Kathryn Stolee, William B. Langdon, and Westley Weimer, editors, *GI-2018, ICSE workshops proceedings*, pages 17–18, Gothenburg, Sweden, 2 June 2018. ACM.

2 Gabin An, Jinhan Kim, and Shin Yoo. Comparing line and AST granularity level for program repair using PyGGI. In Justyna Petke, Kathryn Stolee, William B. Langdon, and Westley Weimer, editors, *GI-2018, ICSE workshops proceedings*, pages 19–26, Gothenburg, Sweden, 2 June 2018. ACM.

3 Benoit Baudry, Nicolas Harrand, Eric Schulte, Marija Selakovic, Shin Hwei Tan, Christopher Timperley, and Emamurho Ugherughe. A spoonful of DevOps helps the GI go down. In Justyna Petke, Kathryn Stolee, William B. Langdon, and Westley Weimer, editors, *GI-2018, ICSE workshops proceedings*, pages 35–37 Gothenburg, Sweden, 2 June 2018. ACM.

4 Joseph Renzullo, Stephanie Forrest, Westley Weimer, and Melanie Moses. Neutrality and epistasis in program space. In Justyna Petke, Kathryn Stolee, William B. Langdon, and Westley Weimer, editors, *GI-2018, ICSE workshops proceedings*, pages 1–8, Gothenburg, Sweden, 2 June 2018. ACM.

## 6.2 New Connections

The followings outcomes were reported by the participants:

1. Eric Schulte, Benoit Baudry, Stephanie Forrest and Nicolas Harrand plan to collaborate on the following topics:
   - The "older but wiser" hypothesis
   - Mapping the "envelope" where executions of neutral variants diverge from one another and identify quiescent points where they converge.
   - Investigating the hypothesis: Systems with more interpretive steps between the "source code" and execution are more robust than those with fewer steps?
2. Eric Schulte, Claire Le Goues and Chris Timperley plan to work at CMU on experimental framework merging (BugZoo)
3. Prof. Banzhaf and Prof. Langdon are planning an experimental evaluation of long term evolution in continuous domains.
4. Dr. Markus Wagner plans to visit Prof. Krawiec Krzysztof during his sabbatical in 2019.
5. Based on a discussion between Prof. Sekanina, Dr. Vasicek and Prof. Krawiec, new research directions have been identified in the area of genetic programming using formal verification methods. Possible ways of collaboration on this topic are under discussion.
6. Dr. Leonardo Trujillo and Dr. John Woodward plan to work on the following question:
   What are the similarities and differences of Decision Forest representations and algorithms and Geometric Semantic Genetic Programming. Both of these approaches have attracted considerable attention over the past few years.
   These two approaches have significant similarities in the types of models they construct, but also some differences.
   They believe these similarities are more than superficial and ask what can these two areas learn from one another.
7. Dr. Claire Le Goues and Prof. Stephanie Forrest will collaborate on round 2 of an idea they tried out several years ago, but now have new ideas for. They will (sometime in the indefinite future) write a paper about improving fitness functions for automated program repair, to include information beyond test suite success.
8. Profs. Colin Johnson and Krzysztof Krawiec talked about the possibilities of using machine learning to measure program quality in genetic improvement and program synthesis, and about the role of program comprehension (and its measurement) in that process. And hope to make progress together in this area.

## Participants


- Brad Alexander
University of Adelaide, AU
- Wolfgang Banzhaf
Michigan State University, US
- Benoit Baudry
KTH Royal Institute of Technology – Stockholm, SE
- Bobby R. Bruce
University College London, GB
- Celso G. Camilo-Junior
Federal University of Goiás, BR
- Myra B. Cohen
University of Nebraska – Lincoln, US
- Benjamin Danglot
INRIA Lille, FR
- Stephanie Forrest
Arizona State University – Tempe, US
- Nicolas Harrand
KTH Royal Institute of Technology – Stockholm, SE
- Colin G. Johnson
University of Kent – Canterbury, GB
- Krzysztof Krawiec
Poznan University of Technology, PL
- William B. Langdon
University College London, GB
- Claire Le Goues
Carnegie Mellon University – Pittsburgh, US
- Alexandru Marginean
University College London, GB
- Michael Pradel
TU Darmstadt, DE
- Joseph Renzullo
Arizona State University – Tempe, US
- Eric Schulte
GrammaTech Inc. – Ithaca, US
- Lukas Sekanina
Brno University of Technology, CZ
- Marija Selakovic
TU Darmstadt, DE
- Shin Hwei Tan
National University of Singapore, SG
- Christopher Timperley
Carnegie Mellon University – Pittsburgh, US
- Leonardo Trujillo
Instituto Tecnológico de Tijuana, MX
- Emamurho Ugherughe
SAP SE – Berlin, DE
- Sergi Valverde
UPF – Barcelona, ES
- Zdenek Vasicek
Brno University of Technology, CZ
- Markus Wagner
University of Adelaide, AU
- John R. Woodward
Queen Mary University of London, GB
- Shin Yoo
KAIST – Daejeon, KR