DAGSTUHL
**REPORTS**

**Volume 13, Issue 9, September 2023**



 ISSN 2192-5283

**ISSN 2192-5283**

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

*Publication date*
March, 2024

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



Digital Object Identifier: 10.4230/DagRep.13.9.i

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

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



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

Report from Dagstuhl Seminar 23361

# Multiobjective Optimization on a Budget

**Richard Allmendinger*[1], Carlos M. Fonseca*[2], Serpil Sayin*[3], Margaret M. Wiecek*[4], and Michael Stiglmayr†[5]**

1 **University of Manchester, GB.** richard.allmendinger@manchester.ac.uk
2 **University of Coimbra, PT.** cmfonsec@dei.uc.pt
3 **Koc University – Istanbul, TR.** ssayin@ku.edu.tr
4 **Clemson University, US.** wmalgor@clemson.edu
5 **Universität Wuppertal, DE.** stiglmayr@uni-wuppertal.de

**Abstract**

The Dagstuhl Seminar 23361 Multiobjective Optimization on a Budget carried on a series of seven previous Dagstuhl Seminars (04461, 06501, 09041, 12041, 15031, 18031, 20031) focused on Multiobjective Optimization. The original goal of this series has been 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. This seminar particularly focused on the case where the approaches from both communities may be challenged by limited resources.

This report documents the program and the outcomes of Dagstuhl Seminar 23361 "Multiobjective Optimization on a Budget". Three major types of resource limitations were highlighted during the seminar: methodological, technical and human related. The effect of these limitations on optimization and decision-making quality, as well as methods to quantify and mitigate this influence, were considered in different working groups.

**Seminar** September 3–8, 2023 – http://www.dagstuhl.de/23361

**2012 ACM Subject Classification** Computing methodologies → Machine learning; Computing methodologies → Neural networks; Mathematics of computing → Evolutionary algorithms; Theory of computation → Mathematical optimization; Computing methodologies → Computational control theory

**Keywords and phrases** evolutionary algorithms, expensive optimization, few-shot learning, machine learning, optimization, simulation

**Digital Object Identifier** 10.4230/DagRep.13.9.1

## 1 Executive Summary

*Richard Allmendinger (University of Manchester, GB)*
*Carlos M. Fonseca (University of Coimbra, PT)*
*Serpil Sayin (Koc University – Istanbul, TR)*
*Margaret M. Wiecek (Clemson University, US)*



Multiobjective optimization (MO), a discipline within systems science that provides models, theories, and methodologies to address decision-making problems under conflicting objectives, has a myriad of applications in all areas of human activity ranging from business and management to engineering. This seminar is a result of the desire to continue to make

* Editor / Organizer
† Editorial Assistant / Collector

MO useful to society as it faces complex decision-making problems and experiences limited resources for decision making. Of particular interest are processes that evolve competitively in environments with scarce resources and lead to decision problems that are characterized by multiple, incommensurate, and conflicting objectives, and engage multiple decision-makers. Viewing optimization and decision making as the complementary aspects of the multiobjective paradigm, the seminar set out to focus around three major types of resource limitations: methodological (e.g., number of solution evaluations), technical (e.g., computation time, energy consumption), and human related (e.g., decision maker availability and responsiveness). The effect of these limitations on optimization and decision-making quality, as well as methods to quantify and mitigate this influence, were of particular interest. Ideas related to modelling, theory, algorithm design, benchmarking, performance metrics, and novel applications of MO under budget constraints were discussed.

To initiate a discussion among the participants on how to address challenges of MO under a budget, the organizers presented specific research directions at the beginning of the seminar. These directions along with their highlights are described below.

- Model reduction: In the MO problem not all functions may be of interest to the decision maker (DM) or not all objectives may be in conflict with each other. Under a limited budget, it is of interest to make the original problem simpler by removing unnecessary objective functions while the solution set remains unchanged. Another reason to reduce the problem is its size. MO problems with four or more criteria bring computational and decision-making challenges that are not typical when the number of objectives is lower.
- Model decomposition and coordination-based decision making: If a reduction of the objectives is not possible, then the solution of the overall MO problem in its entirety may be challenging or even impossible to obtain. In this situation, decomposition of the MO problem into a set of MO subproblems with a smaller number of criteria becomes appealing provided solving the subproblems can be coordinated and related to solving the original problem. When the MO problem is decomposed while computation of the overall solution set is possible, the decomposition goal is to enhance capability of making coordinated tradeoff decisions by working in lower dimensional spaces, which decreases the cognitive burden on DMs. Otherwise, if computation of the overall solution set is not possible, the decomposition goal becomes more challenging since the intention is to coordinate the subproblems' solution sets to construct the overall set and to facilitate decision making in a similar way.
- Representation of the optimization solution set: It is of interest to design cost-effective methods for obtaining a complete or partial description of the Pareto set. An exact description of this set might be available analytically as a closed-form formula, numerically as a set of points, or in mixed form as a parametrized set of points. Unfortunately, for the majority of MO problems, it is not easy to obtain an exact description of the solution set that includes typically a very large number or infinite number of points. Even if it is theoretically possible to find these points exactly, this is often computationally challenging and expensive, and therefore is usually abandoned. On the other hand, if it is possible to obtain the complete solution set, one might not be interested in this task due to overflow of information. Another reason for approximating the solution set, rather than finding the solution set exactly, is that many real-world problems (e.g., in engineering) cannot be completely and correctly formulated before a solution procedure starts. Since the exact solution set is very often not attainable, an approximated description of the solution set becomes an appealing alternative.

- Surrogate-assisted optimization: The combination of evolutionary MO (EMO) algorithms with efficient computational models, often known as metamodels or surrogates, has become a common approach to approximate outcomes of a time-consuming, expensive, and/or resource intense simulation or physical experiment, and thus to tackle problems with a limited budget. Surrogate-assisted (SA) methods vary in aspects such as the use of the metamodel (e.g., different models for different objective functions or one model for all objective functions), type of metamodel (e.g., Gaussian process, radial basis neural network, etc.), how the metamodel is updated (e.g., expected improvement, expected hypervolume improvement), and training time of the metamodel. In particular, the combination of optimization with Gaussian process approximation, known as Bayesian optimization, is a recent trend to efficiently deploy data in model development.
- Multistage optimization: In real-world applications, problem data does not always become available all at once, but at different points in time until a final decision needs to be made. In particular, waiting until all the required data is available may not leave enough time to run the optimization process on the whole problem and successfully compute a final decision. In addition, it is often possible to model the uncertainty associated with the yet unknown data given the data that is already known, at least to some extent. Two-stage (and, more generally, multi-stage) approaches to optimization reformulate the original problem as a number of sub-problems to be solved sequentially, in such a way that the last problem(s) in the sequence can effectively be solved in the (short) time available.
- Preference acquisition and communication with the decision maker: The ultimate goal in MO is to serve one or multiple DMs whose goal is to come up with a single most preferred solution from among the ones that are available. Given an optimization model, DM's preferences may be incorporated prior to, during or after employing a solution procedure. In particular, interactive methods require the DM's involvement in the solution process during which they reveal their preferences based on the presented information. Under a limited budget, communication with the DM shall be designed effectively and economically.
- Benchmarking of algorithms: SA methods are considered as the method of choice to tackle problems subject to a limited budget in terms of function evaluations. However, SA methods are not often compared to widely different alternatives (e.g., different kernels and distance measures, non-SA methods, etc.), and are often tested on narrow sets of problems (multimodal, low-dimensional, static, deterministic, unconstrained, and continuous functions) and rarely on real-world problems, which makes it difficult to assess where (or if) these methods actually achieve state-of-the-art performance in practice. Moreover, several aspects in the design of SA algorithms vary across implementations without a clear recommendation emerging from current practices, and many of these design choices are not backed up by authoritative test campaigns. This seminar topic aimed to raise awareness and hence a push to more work being carried out on developing benchmarking guidelines for SA algorithms.

In response to the presented research directions, some participants found research topics of interest among those suggested by the organizers. These topics included model reduction, decomposition and coordination, solution set representation, and surrogate modeling. Other participants proposed different topics that also targeted the theme of MO under a budget. Those topics included design of experiments for MO, correlations in MO, and design of evolutionary algorithms. Overall, seven research topics were proposed and pursued.

Independently of developing and forming research topics, a collection of eight talks were given during the seminar. Two of the speakers were considered "invited" because they were asked before the seminar to give a talk. These talks addressed two of the research directions initiated by the organizers. The other speakers, being inspired by the ongoing seminar, proposed talks that were integrated daily into the seminar program. The invited and contributed talks kept the seminar in balance ensuring ample time for working in groups.

During the seminar the schedule was updated on a daily basis to maintain flexibility in balancing time slots for the invited and contributed talks, discussions, and working group sessions. The working groups were established on the first day in an interactive fashion. Starting with three large working groups focused around the three central topics of the seminar (methodological, technical, and human-related resource limitations), participants were invited to formulate their favorite topics and most important challenges. The three initial groups split to eventually form eight groups by the end of the seminar. During the week the participants were allowed to change the working groups based on their research interest. The abstracts of the delivered talks and the extended abstracts of the working groups can be found in the subsequent chapters of this report.

Further notable events during the week included: (i) a hike that took place on Wednesday afternoon, (ii) a session allowing the participants to share the details of upcoming professional events in the research community, (iii) a joint session with the participants of the concurrent seminar 23362 “Decision-Making Techniques for Smart Semiconductor Manufacturing” and (iv) an informal get together on Thursday evening.

## Offers and Needs Market

An *Offers* & *Needs Market* ran throughout the entire week. The participants could write their research offers and needs regarding MO on note paper in different colors and post them on pin boards (see Fig. 1) to attract or find a possible collaborator. Participants discussed potential collaboration opportunities during the coffee breaks and after hours.

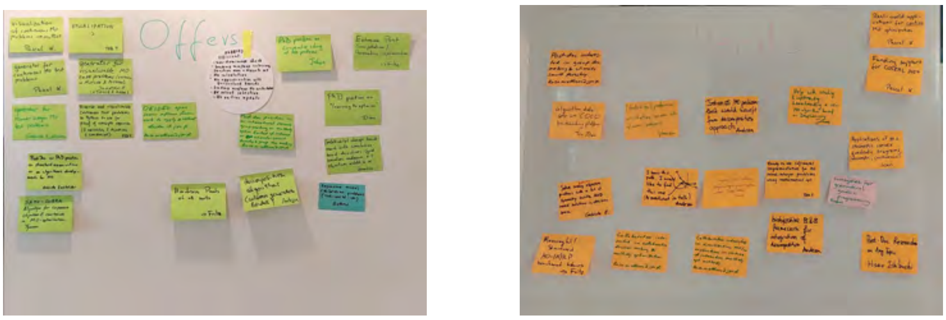



**Figure 1** Offers and needs market.

## Outcomes

The outcomes of each of the working groups can be seen in the sequel.

The organizers have arranged a special issue of the *Journal of Multi-Criteria Decision Analysis* entitled “Multiobjective Optimization on a Budget” for which they will serve as Guest Editors. This issue will be an outlet for papers authored and submitted by the seminar’s participants as well as by researchers world-wide.

This seminar resulted in a very insightful, productive and enjoyable week. It has already led to first new results, cooperations and research topics.

## Acknowledgements

The organizers would like to express their appreciation to the Dagstuhl Office and its helpful and patient staff for their professional support and smooth cooperation; huge thanks to the organizers of the previous seminars in this series for setting us up for success; and thanks to all the participants, who worked hard and were amiable company all week.

In a later section, we also give special thanks to Margaret Wiecek as she steps down from the organizer role.

## 2 Table of Contents

# 3 Overview of Talks

## 3.1 Objective Space Methods: Pareto Front Approximations on a Budget

*Kathrin Klamroth (Universität Wuppertal, DE), Kerstin Dächert (HTW Dresden, DE), Daniel Vanderpooten (University Paris-Dauphine, FR)*


**Joint work of** Kerstin Dächert, Kathrin Klamroth, Renaud Lacour, Daniel Vanderpooten
**Main reference** Kerstin Dächert, Kathrin Klamroth, Renaud Lacour, Daniel Vanderpooten: "Efficient computation of the search region in multi-objective optimization", European Journal of Operational Research, Vol. 260(3), pp. 841–855, 2017.
**URL** https://doi.org//10.1016/j.ejor.2016.05.029

Objective space methods usually rely on (often recursive) decompositions of the objective space, and on the associated formulation of problem scalarizations that are solved by available (single-objective) solvers. The number of solver calls and the complexity of the scalarizations are decisive for the computational effort and may be subject to time, energy or cost constraints. We will briefly review objective space methods and initiate a discussion on the impact of pre-specified budget constraints on the algorithmic choices.

### References

1. K. Dächert. C++ Implementation of the Defining Point Algorithm on github (Version 2023). `https://github.com/kerstindaechert/DefiningPointAlgorithm`
2. K. Dächert and K. Klamroth. A linear bound on the number of scalarizations needed to solve discrete tricriteria optimization problems. *Journal of Global Optimization* 61(4):643-676, 2015
3. K. Dächert, K. Klamroth, R. Lacour and D. Vanderpooten. Efficient computation of the search region in multi-objective optimization. *European Journal of Operational Research* 260(3):841-855, 2017
4. K. Klamroth, R. Lacour and D. Vanderpooten. On the representation of the search region in multi-objective optimization. *European Journal of Operational Research* 245(3):767-778, 2015
5. R. Lacour, K. Klamroth and C.M. Fonseca. A box decomposition algorithm to compute the hypervolume indicator. *Computers & Operations Research* 79:347-360, 2017
6. S. Tamby and D. Vanderpooten. Enumeration of the nondominated set of multiobjective discrete optimization problems. *INFORMS Journal on Computing* 33(1):72-85, 2000

## 3.2 Perspectives to Dealing with Computationally Expensive Multiobjective Optimization Problems

*Kaisa Miettinen (University of Jyväskylä, FI)*



Multiobjective optimization methods are needed since real-life problems typically have several conflicting objective functions to be optimized simultaneously. To find the most preferred Pareto optimal solution as the final one to be implemented in practice, we typically need preference information from a domain expert, a decision maker (DM).

We concentrate on interactive methods, where the DM takes actively part and directs the solution process with one's preference information. One can learn and gain insight about the problem and also adjust preferences while learning. Importantly, one can concentrate on solutions that seem most promising and avoid high cognitive load of analyzing too much information at a time.

In real applications, function evaluations may be expensive and we outline different approaches. The first is to generate a representative set of Pareto optimal solutions in advance and create a surrogate problem that is computationally inexpensive. In the second approach, we replace a scalarizing function that a multiobjective optimization method employs by a computationally inexpensive metamodel. The third approach is to fit a metamodel to each computationally expensive objective function. Appropriate approaches are also needed if constraint functions are expensive or functions in the problem to be solved have different latencies.

By speeding up calculations, we avoid keeping the DM waiting when applying interactive methods. But the presence of the human DM means that attention must be paid to the understandability and amount of information expected from the DM. We briefly outline pros and cons of some methods and mention further challenges.

## 3.3 Surrogate model guided optimization of expensive black-box multiobjective problems

*Juliane Mueller (NREL – Golden, US)*



Many engineering applications require the simultaneous optimization of multiple conflicting objective functions. Often, these objective functions are evaluated using highly accurate computer simulations that are computationally too expensive to be evaluated hundreds or thousands of times during optimization. Thus, the goal is to find good approximations of the Pareto front using as few of these expensive simulations as possible. Here, we describe an optimization approach based on surrogate models and diverse sampling strategies to accelerate the search for the Pareto solutions. We use a separate surrogate model for approximating each objective function and then we use the surrogate models to inform where additional expensive simulations should be run. The surrogate models are updated in an active learning framework whenever new information from the expensive simulations becomes available. The sampling strategies aim at balancing local improvements of the approximate Pareto front and global exploration to identify the extrema and fill in large gaps of the approximate Pareto front. We demonstrate on a large set of benchmark problems the effectiveness of the method for finding good approximations of the Pareto front.

## 3.4 Fast Pareto Optimization Using Sliding Window Selection

*Frank Neumann (University of Adelaide, AU)*


**Joint work of** Frank Neumann, Carsten Witt
**Main reference** Frank Neumann, Carsten Witt: "Fast Pareto Optimization Using Sliding Window Selection", in Proc. of the ECAI 2023 – 26th European Conference on Artificial Intelligence, September 30 – October 4, 2023, Kraków, Poland – Including 12th Conference on Prestigious Applications of Intelligent Systems (PAIS 2023), Frontiers in Artificial Intelligence and Applications, Vol. 372, pp. 1771–1778, IOS Press, 2023.
**URL** https://doi.org//10.3233/FAIA230463

Pareto optimization using evolutionary multi-objective algorithms such as the classical GSEMO algorithm has been widely applied to solve constrained submodular optimization problems. A crucial factor determining the runtime of the used evolutionary algorithms to obtain good approximations is the population size of the algorithms which grows with the number of trade-offs that the algorithms encounter. In this paper, we introduce a sliding window speed up technique for recently introduced algorithms. We prove that our technique eliminates the population size as a crucial factor negatively impacting the runtime of the classical GSEMO algorithm and achieves the same theoretical performance guarantees as previous approaches within less computation time. Our experimental investigations for the classical maximum coverage problem confirms that our sliding window technique clearly leads to better results for a wide range of instances and constraint settings.

## 3.5 Towards decision analytic workflows for real-world problems: Simulation model calibration and multi-objective optimization on a shared evaluation budget

*Robin Purshouse (University of Sheffield, GB)*


**Joint work of** Oliver P. H. Jones, Jeremy E. Oakley, Robin C. Purshouse
**Main reference** Oliver P. H. Jones, Jeremy E. Oakley, Robin C. Purshouse: "Simulation-based engineering design: solving parameter inference and multi-objective optimization problems on a shared simulation budget", in Proc. of the 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1399–1405, 2021.
**URL** https://doi.org//10.1109/SMC52423.2021.9658645

Real-world multi-objective optimization problems sometimes include evaluation functions that rely on computationally expensive simulation models. These types of problems typically constrain the optimization budget to a relatively small number of candidate solutions, e.g. 500-5000. An often-overlooked issue in such problems is that the simulations (i.e. evaluation functions) typically require the calibration of their parameters before they are ready for use in solving a particular problem instance. The simulations can also contain discrepancies – e.g. simplifications in the representation of the physics of the problem–that affect the robustness of solution evaluations. Simulation model calibration is a research field in its own right and concerns itself with inference of model parameters and model discrepancy structures. Inference is typically computational in nature, uses Bayesian methods, and involves the evaluation of sampled candidate parameterisations and discrepancy terms via the simulation model – i.e. it also involves evaluation functions and a constrained evaluation budget. To improve the efficiency of the inference process, low fidelity "emulators", often

Gaussian processes, are estimated using data sampled from the high fidelity model, closely mirroring the use of surrogate models in optimization workflows. Perhaps surprisingly, very little research has been conducted into the joint problem of simulation model calibration and multi-objective optimization based on such models. How should an evaluation budget best be allocated to the two activities, how should they be sequenced, and how can synergies between the two be exploited? This presentation introduces this novel topic, demonstrates some illustrative benchmark problems, and sketches some tentative solution architectures.

Acknowledgements: Robin Purshouse is supported by SIPHER (MR/S037578/1), a UK Prevention Research Partnership funded by the UK Research and Innovation Councils, the Department of Health and Social Care (England) and the UK devolved administrations, and leading health research charities https://ukprp.org/. Oliver Jones was supported by the UK Engineering and Physical Sciences Research Council.

### References

1 Jones OPH, Oakley JE, Purshouse RC, Toward a unified framework for model calibration and optimisation in virtual engineering workflows, 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2019, pp. 3148–3153.
2 Jones O, A framework for combining model calibration with model-based optimization in virtual engineering design workflows, PhD thesis, University of Sheffield, 2021.
3 Jones OPH, Oakley JE, Purshouse RC, Simulation-based engineering design: solving parameter inference and multi-objective optimization problems on a shared simulation budget, 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2021, pp. 1399–1405.

## 3.6 Efficient Approximation of Expected Hypervolume Improvement Using Gauss-Hermite Quadrature

*Alma Rahat (Swansea University, GB)*


**Joint work of** Alma A. M. Rahat, Tinkle Chugh, Jonathan E. Fieldsend, Richard Allmendinger, Kaisa Miettinen
**Main reference** Alma A. M. Rahat, Tinkle Chugh, Jonathan E. Fieldsend, Richard Allmendinger, Kaisa Miettinen: "Efficient Approximation of Expected Hypervolume Improvement Using Gauss-Hermite Quadrature", in Proc. of the Parallel Problem Solving from Nature – PPSN XVII – 17th International Conference, PPSN 2022, Dortmund, Germany, September 10-14, 2022, Proceedings, Part I, Lecture Notes in Computer Science, Vol. 13398, pp. 90–103, Springer, 2022.
**URL** https://doi.org//10.1007/978-3-031-14714-2_7

Many methods for performing multi-objective optimisation of computationally expensive problems have been proposed recently. Typically, a probabilistic surrogate for each objective is constructed from an initial dataset. The surrogates can then be used to produce predictive densities in the objective space for any solution. Using the predictive densities, we can compute the expected hypervolume improvement (EHVI) due to a solution. Maximising the EHVI, we can locate the most promising solution that may be expensively evaluated next. There are closed-form expressions for computing the EHVI, integrating over the multivariate predictive densities. However, they require partitioning of the objective space, which can be prohibitively expensive for more than three objectives. Furthermore, there are no closed-form expressions for a problem where the predictive densities are dependent, capturing the correlations between objectives. Monte Carlo approximation is used instead in such cases, which is not cheap. Hence, the need to develop new accurate but cheaper approximation methods remains. Here we investigate an alternative approach toward approximating the

EHVI using Gauss-Hermite quadrature. We show that it can be an accurate alternative to Monte Carlo for both independent and correlated predictive densities with statistically significant rank correlations for a range of popular test problems.

## 3.7 Problem decomposition in biobjective optimisation

*Andrea Raith (University of Auckland, NZ)*


**Joint work of** Matthias Ehrgott, Ruchard Lusby, Andrew Mason, Siamak Moradi, Melanie Reuter-Oppermann, Ali Sohrabi Yousefkhan

Decomposition techniques for optimisation problems have significantly improved the ability to solve problems of ever-increasing complexity and problem size by decomposing a complex optimisation problem into related smaller ones. The premise of a decomposition technique is to omit parts of the problem that are unlikely to influence the final solution, and iteratively include, as needed, the parts which will have an impact. Many real-world problems must be formulated with two or more objectives and solving such multiobjective optimisation problems means identifying sets of so-called efficient solutions representing available trade-offs. Different solution algorithms for biobjective linear programmes (BLPs) will be discussed in this talk. Building on a biobjective (parametric) version of the well-known simplex algorithm, different decomposition approaches are presented here. One approach, also known as column generation, is to omit some or all variables (corresponding to columns of the constraint matrix) from the original optimisation problem and then iteratively re-introduce them into the problem. An alternative approach, known as Benders decomposition, separates decision variables into different stages and related optimisation problems, and then dynamically adds constraints into the first-stage formulation to capture the full problem. We present theoretical developments and algorithms that adapt these ideas into decomposition techniques for BLPs. We will also briefly discuss initial developments of a so-called math-heuristic approach that combines exact optimisation concepts with a neighbourhood search heuristic that can be used instead of an exact column generation approach.

## 3.8 A Visualization-Aided Approach to Solving Tri-Criterion Portfolio Problems

*Ralph E. Steuer (University of Georgia – Athens, US)*


**Joint work of** Ralph E. Steuer, Sebastian Utz

This talk contains no text or equations, only graphs. It is about (1) how much more enormously difficult it is to identify one's best point on an efficient surface than on an efficient frontier and (2) how, on problems in which one objective appears to carry more influence, a visually assisted approach utilizing a new type of line stretched across the efficient set can be applied. Of course, the approach works best the less severely disordered the efficient surface is. A non-trivial tri-criterion portfolio optimization problem is used to illustrate throughout.

## 3.9 Multi-objective Branch-and-Bound on a Budget

*Michael Stiglmayr (Universität Wuppertal, DE)*


**Joint work of** Julius Bauß, Michael Stiglmayr

In this talk we discuss modifications of multi-objective branch-and-bound to diversify solutions and yield a good approximation of the non-dominated set when only limited computation time is available. It is crucial not only to find efficient solutions in early stages of the algorithm but also to find a set of solutions whose images are close to and well distributed along the non-dominated frontier. In particular the adaptation of branching and queuing of sub-problems seems to be important. We use, e.g., the hypervolume indicator as a measure for the gap between lower and upper bound set to implement a multi-objective best-first strategy. Moreover, gap measure indicate the solution quality when prematurely stopping the branch-and-bound algorithm.

### References

1 Bauß, J., Stiglmayr, M.: Augmenting bi-objective branch and bound by scalarization-based information (2023), https://arxiv.org/abs/2301.11974
2 Bauß, J., Stiglmayr, M.: Adapting branching and queuing for multi-objective branch and bound (2023), https://arxiv.org/abs/2311.05980

# 4 Working groups

## 4.1 Decoupled Design of Experiments for Multi-objective Optimisation on a Budget

*Mickaël Binois (INRIA – Sophia Antipolis, FR), Jürgen Branke (University of Warwick, GB), Jonathan Fieldsend (University of Exeter, GB), Robin Purshouse (University of Sheffield, GB)*



### 4.1.1 Introduction

Fundamental to the performance of surrogate-based optimisation frameworks is the need to construct an *initial model* based on a carefully selected set of initial designs, and any prior system knowledge. This is both in the case of Bayesian optimisation, which used and iteratively update model(s) mapping decision vectors to predicted performance criteria values, and for evolutionary computation approaches which involving surrogates. The selection and construction of initial designs, which are often treated separately to the decision vectors queried during the subsequent optimisation process, are usually referred to as the *design of experiments* (or DoE for short). This is because these decision vectors are selected to – in some fashion – be maximally informative on the global underlying process, rather than being biased towards particular regions.

Without any prior information regarding the properties of the objective function(s) such DoE for model fitting are commonly based around *space filling* sequences such as Latin hypercube sampling [9] or Sobol sequences [10], as purely random sampling tends to naturally result in clusters, which do not serve model fitting well, particularly when the budget for sampling is tight.

Where there are multiple criteria being modelled, this leads to an interesting and under-explored question: *should one evaluate all initial designs fully, or selectively evaluate a subset of objectives per design, allowing a greater number of locations to be partially evaluated when building the model(s)?* A few works have looked at *decoupling* objective evaluations during the search process – particularly where there are different costs associated with each objective, but this can also be advantageous where there is a difference in the complexity of the functions being modelled (e.g. one being smooth slowly changing, the other being rugged and fast changing). As such, this appears to be a promising direction for further investigation and research, as even small improvements in such areas can effectively lead to large savings for expensive optimisation problems.

### 4.1.2 Related Work

A small number of existing works have considered decoupled and/or cost-aware multi-objective optimisation – some of which have considered these factors during the initial DoE phase. Below we discuss the most relevant approaches. A wider survey on the topic of objectives with different costs can be found in [1].

Hernández-Lobato and colleagues proposed the *Predictive Entropy Search for Multi-Objective Bayesian Optimization* (PESMO) method [6]. PESMO uses predictive entropy search as the acquisition function. This function represents each objective using an additive component, which enables a decoupled evaluation approach to be adopted. The approach was subsequently extended to also consider constraints (again where decoupling is possible) [5].

Suzuki et al. developed the *Pareto-frontier entropy search* (PFES) approach [11]. PFES is also an entropy approach but considers the entropy in objective-space rather than decision-space, which is computationally simpler. This method also includes cost in evaluating the objectives by including cost in the denominator of the acquisition function. Like PESMO, the approach is easily extended to consider decoupled evaluations.

Iqbal and colleagues proposed the *Flexible Multi-Objective Bayesian Optimization* (FlexiBo) algorithm [7]. The approach uses a decoupled evaluation in the Bayesian optimisation run but uses a coupled initial DoE procedure. FlexiBo includes two main features: (1) a new acquisition function that is the expected change in hypervolume if only one objective function is evaluated, divided by the cost of this function evaluation; and (2) a confidence region in the objective space for the partially evaluated points. The estimated cost of evaluating each objective is updated each time the objective is evaluated – this is a mean estimate of the cost (treating any observed variability as occurring at random).

Most recently, Buckingham et al. extended the multi-attribute Knowledge Gradient [2] to the case where objectives can be evaluated independently [3]. The authors demonstrate the benefit of independent evaluation not only when the computational times for objectives differ, but also when the lengthscales of the modelled landscapes (which determine the smoothness of the landscape) differ.

A slightly different problem is considered in [8], where one objective is much cheaper (essentially free) to evaluate than the other. They directly incorporate evaluation of the cheap objectives into a pair of hypervolume-based acquisition functions for BO. Consequently, the cheap objectives are evaluated many times while the acquisition function is optimized.

A summary of the different approaches is shown in Table 1, highlighting which methods feature decoupled and cost-aware acquisition functions during the initial DoE, the subsequent optimisation run, or both phases.

**Table 1** Existing methods for decoupled cost-aware multi-objective optimisation.

| | Design of experiments | | Optimisation | | |
|---|---|---|---|---|---|
| Approach | Decoupled? | Cost-aware? | Decoupled? | Cost-aware? | Acquisition function |
| PESMO [6] | ✓ | ✗ | ✓ | ✗ | predictive entropy search |
| PFES [11] | ✗ | ✗ | ✓ | ✓ | cost-weighted Pareto frontier entropy |
| FlexiBO [7] | ✗ | ✗ | ✓ | ✓ | cost-weighted objective space entropy |
| C-MOKG [3] | ✗ | ✗ | ✓ | ✓ | cost-weighted multi-objective knowledge gradient |

### 4.1.3 New analyses by the working group

#### 4.1.3.1 Initial DoE when evaluations are decoupled

The costs of the objectives are assumed to be the same for now.

Goal: studying the effect on coupled vs. decoupled designs of experiments (DoE) on the uncertainty on the Pareto front.

To this end, we experiment on Gaussian process models (GPs). More precisely, we generate samples from a Gaussian process model and use it as the ground truth. The hyperparameters are supposed to be known to remove the effect of inference. Hence there is no model mismatch. Examples of outcome are given in Figure 2.

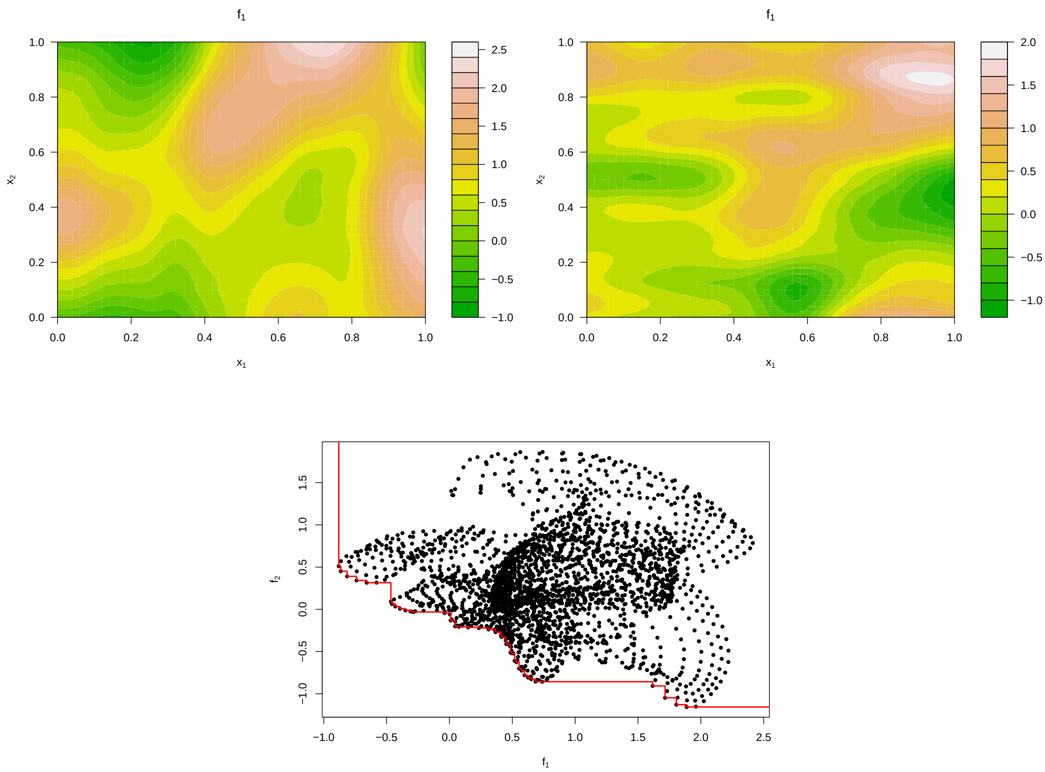



**Figure 2** Top: two realisations of Gaussian process priors, with Matérn 5/2 covariance kernel, with lengthscale hyperparameters $(0.3, 0.4)$ (resp. $(0.4, 0.2)$) for $f_1$ (resp. $f_2$), and unit variance. Bottom: corresponding image in the objective space.

Next, to measure the uncertainty on the Pareto front associated with the fitted GPs, we rely on the so called Vorob'ev deviation (VD), a set based variance measure, see Algorithm 1 for a pseudo code and, e.g., [4] for the details. The reference point used for hypervolume

**Algorithm 1:** Pseudo-code for the testing procedure.

Generate the first design of experiments $X_1$ for objective 1.
(Coupled case) $X_2 = X_1$ the DoE of the second objective is the same.
(Decoupled case) Generate $X_2$ the second DoE.
Build GP models.
Generate $s$ conditional samples on some designs $X_s$ from all GPs.
Compute the $s$ non-dominated points on couples of samples from the different GPs.
Compute the corresponding Vorob'ev deviation.

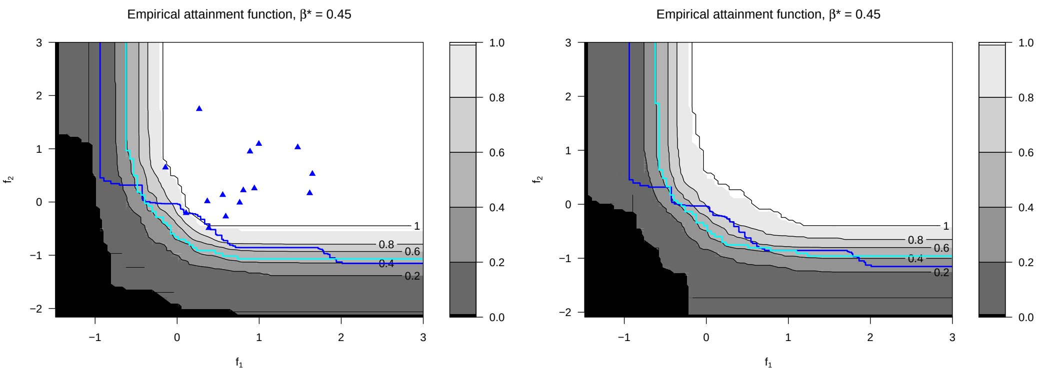



**Figure 3** Attainment function representation in the coupled (left) and decoupled (right) cases. The blue triangle mark observations in the coupled case, where both objectives are evaluated. The cyan line represents the estimated Pareto front of the GP while the reference Pareto front is in blue.

computations is taken to be $(3, 3)$. An example is provided in Figure 3, where the DoE for the first objective is the same while the second one is either coupled or decoupled. One visible effect is that when both objectives are jointly evaluated, the area that is dominated (attainment value $= 1$) is larger. This is probably because in the decoupled case, solutions are never surely dominated (even though the domination probability is extremely low).

We compare VD values of different setups for the coupled and decoupled case:

- the DoE for the first objective is either uniformly sampled or is a maximin Latin hypercube design;
- the DoE for the second objective is the same as the first objective (coupled case), uniformly sampled or an LHS augmenting the DoE of the first objective.

Figure 4 shows the results. Comparing the top row (both initial designs uniformly sampled) with the middle row (an augmenting LHS used to complement the first uniform DoE), there seems to be not much difference. However, the bottom row (first design is sampled with LHS, second uses augmenting LHS) shows a significant improvement of the Vorob'ev deviation of either the coupled (red dots) or decoupled (box plots) sampling. Clearly, a space-filling design improves our estimate of the Pareto front, but it seems not sufficient to only make the design of the second objective space-filling.

Note that with respect to the Vorob'ev deviation, when at least one of the designs is random (first two rows, first two columns), the red dots are sometimes above and sometimes below the median of the boxplots, while the red dots are mostly below the median of the boxplots in the bottom row (full space-filling design). This indicates that at least if a space-filling design is used, decoupled sampling is worse than coupled sampling, possibly due to the effect mentioned above on the size of the known dominated region. Note, however, that in these experiments we assume equal cost of sampling the two objectives, and equal lengthscales of the two objectives. As we see later, in other cases decoupling may be beneficial.

The results look slightly different when considering the expected product of the standard deviations of the GP (right column), which is an indication of the accuracy of the estimation quality of the models over the entire search space, rather than the Pareto frontier. Here, the first two rows show a clear benefit of decoupled sampling. However, this benefit seems to disappear once both objectives are sampled using space-filling designs (third row).

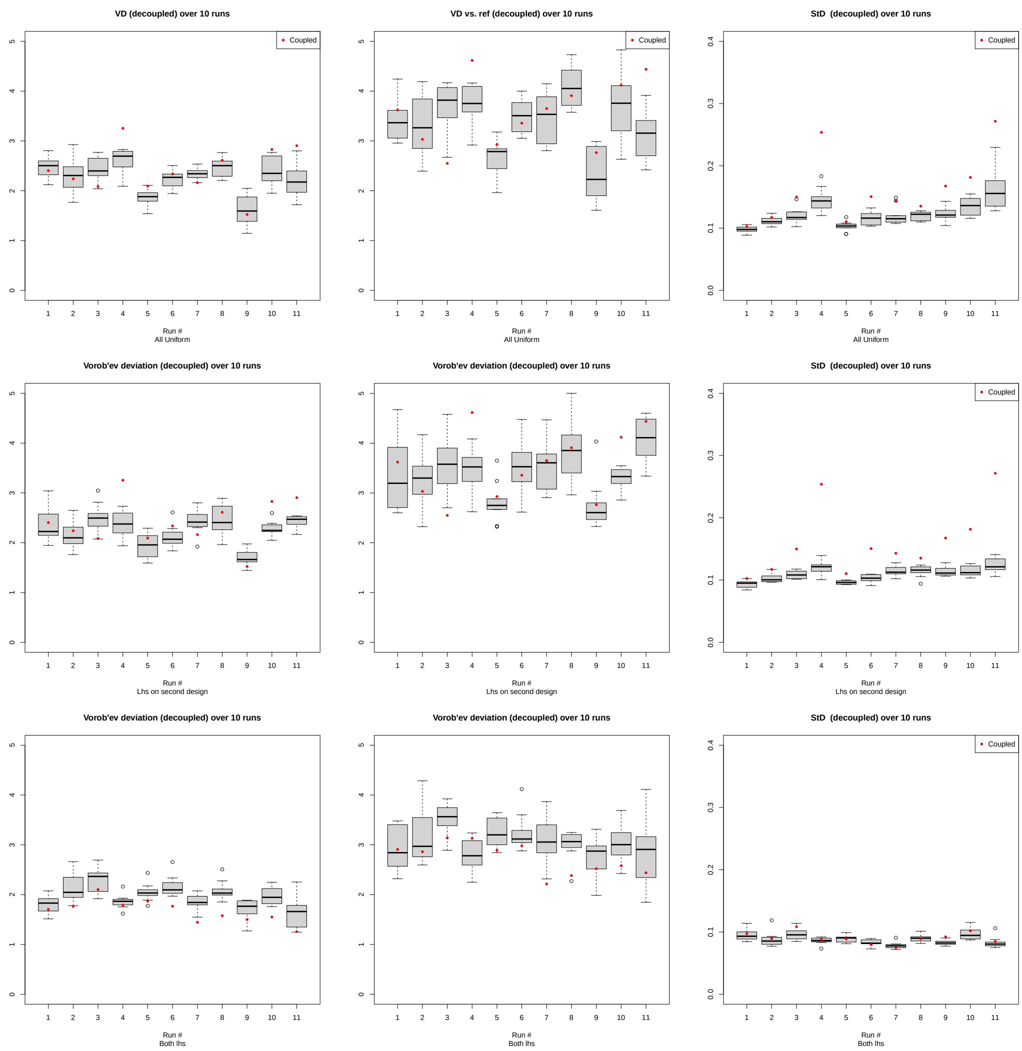



**Figure 4** Boxplots of Vorob'ev deviation with decoupled designs, over 11 different runs and 10 replications per run. In the top row, both initial designs are uniformly sampled, in the middle row, an augmenting LHS is used to complement the first uniform DoE, and in the bottom row, an augmenting LHS is used to complement the first LHS design. Left column shows VD, middle column shows VD against true Pareto front, and right column shows standard deviation product. The value of the coupled design is in red.

#### 4.1.3.2 Initial DoE when evaluations have different costs

Now let us assume the cost is different between different objectives $f_1$ and $f_2$ (etc). The first tasks are to define the total time budget for experiments and get relative costs of $f_1$, $f_2$, ..., $f_3$. We will then consider a number of alternative approaches to DoE, including a coupled baseline.

1. (Coupled) Both functions evaluated at once.
2. (Decoupled naive) Both functions evaluated the same number of times, but at differing locations. (generated by Augmented LHS)
3. (Decoupled) The allocation of total budget to the two functions depends on lengthscales and relative costs, according to Eq. 1. Objectives with smaller lengthscales and smaller cost are sampled more often.

Considering how to split the computational budget, let us consider the simplest case of optimising a (weighted) sum of two objectives. In such a case, if we want to minimise integrated mean squared prediction error (IMSPE), then it is not possible to improve beyond coupled sampling, as the variances of the two functions just add up, and the optimal design for each function would be the same. However, if the costs or lengthscales are different, then we could use IMSPE to determine an appropriate allocation of the budget to the two functions as follows:

$$\min \frac{IMSPE(n_1)}{c_1 \times n_1} + \frac{IMSPE(N - n_1)}{c_2 \times (N - n_1)}, \tag{1}$$

where $N$ is the total budget, $n_1$ is the number of samples allocated to objective $f_1$, and $c_1(c_2)$ are the cost of evaluating objective $f_1(f_2)$.

As in the previous section, we rely on GP samples to define a ground truth. We also assume some known values of the lengthscales of the objectives: $(0.3, 0.4)$ for the first, $(0.4, 0.2)$ for the second. We start with four initial designs for each objective in the various cases, then 26 decoupled evaluations are performed. We only compare the 'coupled', 'naive' and 'decoupled' strategies. The results are in Figure 5. First, from the IMSPE results, we observe that the values for objective 1 and 2 are different (importantly, the GP variances are equal here), due to the different lengthscales. The naive baseline always performs worst. Then, in the same cost case, there is no change between the coupled and decoupled case. As the cost of $f_2$ increases, the effect is that the IMSPE of $f_1$ is reduced faster compared to $f_2$, with no strong detrimental effect on $f_2$ for the same total cost. The outcome is that it is reasonable to sample more $f_1$, in a ratio that only depends on the lengthscales and relative cost.

### 4.1.4 Discussion and future research ideas

In this report, we have examined the possibility of improving the quality of the surrogate models obtained through a DoE in case of multi-objective optimisation where the evaluation of the different objectives can be decoupled. We found that for the case of equal lengthscales, decoupling the evaluations (i.e., evaluating different solutions on different objectives) did tend to worsen the quality of the Pareto front estimate as measured by Vorob'ev deviation. However, when objectives had different costs and/or lengthscales, decoupling could improve results substantially in terms of total IMPSE.

In the future, we plan to investigate also other sampling strategies such as taking into account the posterior of the first objective when deciding where to evaluate the second objective, or to learn each objective function's lengthscale and cost on the fly.

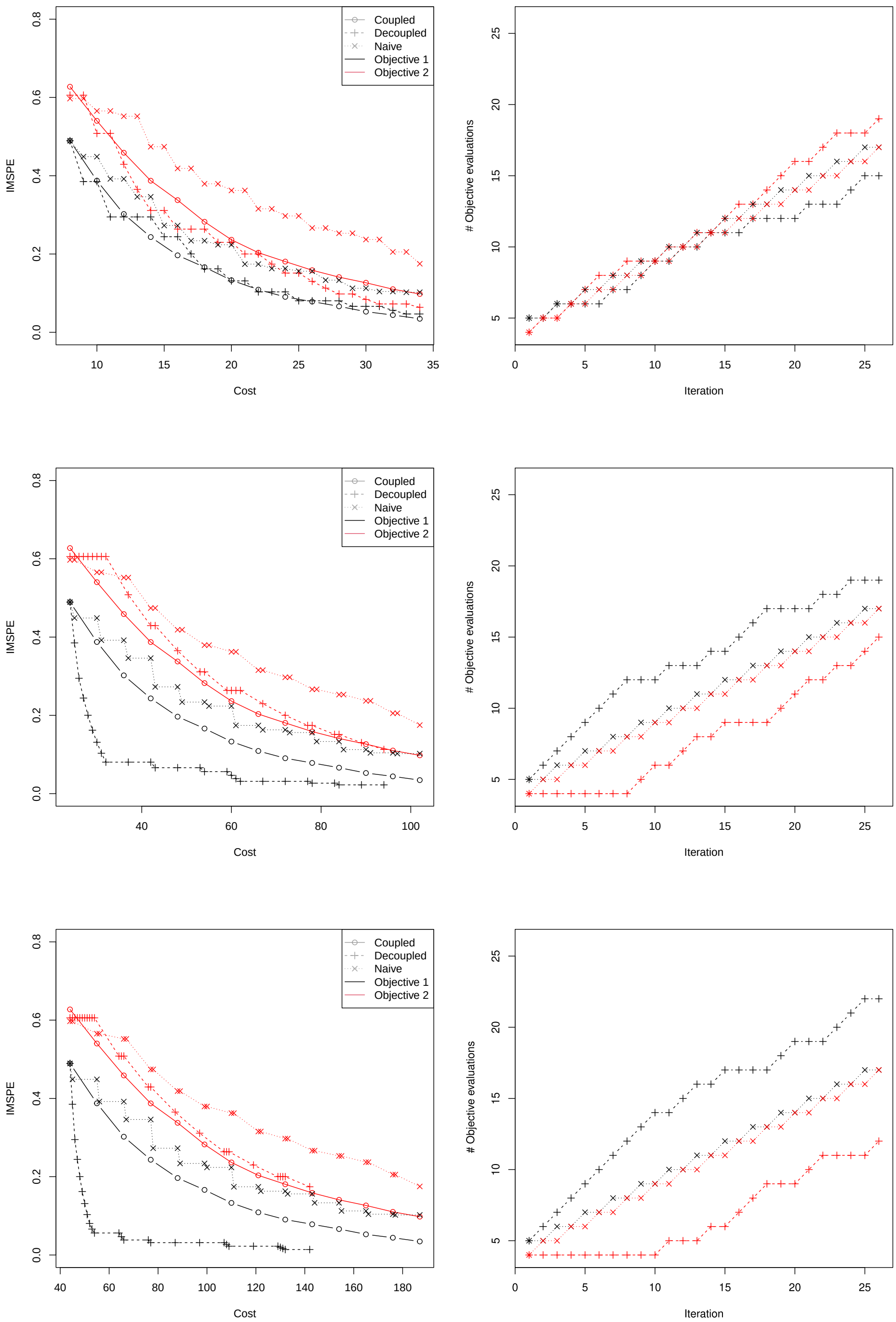



**Figure 5** Left: IMSPE vs. cost for the various strategies. Right: objectives evaluated per iteration. Top: cost is equal for both objectives, Middle: cost of $f_2$ is 5 times greater, Bottom: cost of $f_2$ is 10 times greater.

### 4.1.5 Acknowledgements

This report benefited from wider discussions within the "Surrogates" working group of the Dagstuhl Seminar 23361 Multiobjective Optimization on a Budget. This group's members included Thomas Bäck, Mickaël Binois, Jürgen Branke, Jonathan Fieldsend, Ekhine Irurozki, Pascal Kerschke, Boris Naujoks, Robin Purshouse, Tea Tusar, Vanessa Volz, Hao Wang and Kaifeng Yang.

**References**

1 Richard Allmendinger and Joshua Knowles, Heterogeneous Objectives: State-of-the-Art and Future Research, *arXiv preprint arXiv:2103.15546*, 2, 2021
2 Raul Astudillo and Peter Frazier, Multi-attribute Bayesian optimization under utility uncertainty In: *Proceedings of the NIPS Workshop on Bayesian Optimization*, 172, 2017.
3 Jack M. Buckingham, Sebastian Rojas Gonzalez and Juergen Branke, Bayesian Optimization of Multiple Objectives with Different Latencies, *arXiv preprint arXiv:2302.01310*, 2023.
4 Mickaël Binois, David Ginsbourger and Olivier Roustant, Quantifying uncertainty on Pareto fronts with Gaussian process conditional simulations, *European Journal of Operational Research*, 243(2), 386–394, 2015.
5 Eduardo C Garrido-Merchán and Daniel Hernández-Lobato, Predictive Entropy Search for Multi-objective Bayesian Optimization with Constraints, *Neurocomputing*, 361, 50–68, 2019.
6 Daniel Hernández-Lobato, Jose Hernandez-Lobato, Amar Shah and Ryan Adams, Predictive entropy search for multi-objective Bayesian optimization, In: *International Conference on Machine Learning*, 1492–1501, PMLR, 2016.
7 Md Shahriar Iqbal, Jianhai Su, Lars Kotthoff and Pooyan Jamshidi, FlexiBO: A Decoupled Cost-Aware Multi-Objective Optimization Approach for Deep Neural Networks, *Journal of Artificial Intelligence Research*, 77, 645–682, 2023.
8 Nasrulloh Loka, Ivo Couckuyt, Federico Garbuglia, Domenico Spina, Inneke Van Nieuwenhuyse and Tom Dhaene, Bi-objective Bayesian optimization of engineering problems with cheap and expensive cost functions, *Engineering with Computers*, 1, 2022.
9 M. D. Mckay, R. J. Beckman and W. J. Conover, A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code, *Technometrics*, 42(1), 55–61, 2000.
10 Harald Niederreiter, *Random number generation and quasi-Monte Carlo methods*, SIAM, 1992.
11 Shinya Suzuki, Shion Takeno, Tomoyuki Tamura, Kazuki Shitara and Masayuki Karasuyama, Multi-objective Bayesian optimization using Pareto-frontier entropy, In: *International Conference on Machine Learning*, 9279–9288, PMLR, 2020.

## 4.2 Hypervolume-Indicator-Based Evolutionary Algorithms on a Budget

*Jürgen Branke (University of Warwick, GB), Kerstin Dächert (HTW Dresden, DE), Andrzej Jaszkiewicz (Poznan University of Technology, PL), Kathrin Klamroth (Universität Wuppertal, DE)*



### 4.2.1 Motivation

Indicator-based evolutionary algorithms are among the most powerful multi-objective algorithms, in particular when using hypervolume (HV) contribution as indicator. They are not really suitable for many-objective problems, as the computational cost for computing HV contributions becomes prohibitive as the number of objectives increases. This working group was looking at ways to make HV-indicator based evolutionary algorithms computationally feasible also for many-objective problems. Since steady-state indicator-based EAs in every iteration only generate one new solution and remove one solution from the population, we saw the following opportunities:

1. Identifying the solution with the least HV contribution may be easier than computing the HV contribution for all solutions.
2. The problems to be solved in each iteration are very similar, with only one solution replaced by another. A lot of the computational results may be transferred from one iteration to the next.
3. Evolutionary algorithms are stochastic algorithms and inherently tolerant to noise. Thus it is not clear whether it is actually necessary to always correctly identify the solution with the minimal HV contribution. Perhaps an approximation with tuneable precision would be sufficient and not jeopardise the optimisation behaviour of the MOEA.
4. HV calculations, in particular Monte Carlo approximations, could benefit from the use of GPUs.

### 4.2.2 Basic definitions

▶ **Definition 1.** *Hypercuboid-bounded hypervolume indicator.* Given a set of points $S$ in the objective space $\mathbb{R}^d$ and a hypercuboid $[r_*, r^*]$ such that $\forall_{p \in S}\, r_* \preceq p$, the hypercuboid-bounded hypervolume indicator of S is the measure of the region weakly dominated by $S$ within $[r_*, r^*]$, i.e.:

$$H(S, [r_*, r^*]) = \mathcal{L}(\{q \in [r_*, r^*] \mid \exists p \in S : q \preceq p\}) \tag{2}$$

where $\mathcal{L}(.)$ denotes the Lebesgue measure.

Note that

$$H(S, [r_*, r^*]) = H(\text{nd-worse}(S, r^*), r_*) \tag{3}$$

where $\text{nd-worse}(S, r^*) = \{q \in \mathbb{R}^d \mid \exists q' \in S : \forall j\, q_j = \min(q'_j, r^*_j)\}$ may be interpreted as projection of $S$ onto $[r_*, r^*]$. Note also that Hypercuboid-bounded hypervolume is equivalent to the standard hypervolume if all points in $S$ are weakly dominated by $r^*$.

▶ **Definition 2.** *Hypervolume contribution.* Hypervolume contribution of a point $s$ to $H(S \cup \{s\}, r_*)$ (allowing both $s \in S$ or $s \notin S$) is the difference between hypervolume of $S \cup \{s\}$ and hypervolume of $S \setminus \{s\}$, i.e.:

$$HVC(s, S, r_*) = HV(S \cup \{s\}, r_*) - HV(S \setminus \{s\}, r_*) \tag{4}$$

Hypervolume contribution of a point $s$ defined by equation (4) could alternatively be calculated as the difference of hypervolume of $\{s\}$ and hypercuboid-bounded hypervolume of $S \setminus \{s\}$ within $[r_*, s]$, i.e.:

$$HVC(s, S, r_*) = \mathcal{L}([r_*, s]) - HV(S \setminus \{s\}, [r_*, s]) \tag{5}$$

where $\mathcal{L}([r_*, s])$ is the hypervolume of hypercuboid $[r_*, s]$. In practice, the use of equation (5) allows for a faster calculation of hypervolume contribution than equation (4), since hypervolume is calculated just once (the time of calculation of $\mathcal{L}([r_*, s])$ is negligible) and many points in $S$ may become dominated after projection onto $[r_*, s]$.

### 4.2.3 Literature review

There is a lot of literature on how to efficiently compute the HV, and also some literature on how to compute HV contributions. In fact, it is too much to list here, so the interested reader is instead referred to the surveys in [6, 11].

[8] proposed the Quick Extreme Hypervolume Contribution (QEHC) algorithm which may be used to efficiently find a point (solution) with either the minimum or the maximum contribution for any number of objectives. Since calculation of hypervolume contribution boils down to the calculation of hypervolume (see (5)), QEHC uses the algorithm introduced in [7] to calculate concurrently HV contribution of each solution with quickly converging guaranteed lower and upper bounds for hypervolume contribution. It then uses these bounds to stop calculation of HV contributions for solutions that may not give the unique minimum or maximum contribution.

[3] and [9] proposed greedy lazy approaches for hypervolume subset selection problem with either incremental or decremental approach. These algorithms in each iteration select the solution with the maximum ([3]) or minimum ([9]) hypervolume contribution and utilize the fact that hypervolume is a non-decreasing submodular function:

$$HVC(s, S \cup \{p\}, r_*) \leq HVC(s, S, r_*), \forall S, \forall p \notin S \tag{6}$$

Since these algorithms only add ([3]) or remove ([9]) a solution in each iteration, they may use contributions calculated in a previous iteration as the upper or lower bounds in a subsequent iteration. Note, however, that these bounds cease to be valid if solutions are both added and removed form a set of solutions.

[4] extends [3] and exploits submodular properties of the HV indicator to reduce the number of HV contribution calculations when selecting a subset from a large number of Pareto-optimal solutions. [10] proposes a local search method for selecting the subset of size $k$ with maximum HV from a larger set of Pareto optimal solutions. Among other things, they show that the number of solutions whose HV contribution is affected by removal of one solution grows very quickly with increasing number of objectives (Fig. 4 in [10]).

[2] propose a fast approximation algorithm to determine the solution with the smallest HV contribution. For given $\epsilon, \delta > 0$ it identifies, with probability at least $(1 - \delta)$, a solution with contribution at most $(1 + \epsilon)$ times the true minimal HV contribution. It is shown to work on very large problem instances with thousands of solutions and hundred dimensions.

[12] proposes a method to efficiently approximate a solution's HV contribution using line segments.

[5] develop a neighborhood structure among local nadir points (also referred to as local upper bounds) in order to compute the entire nondominated set of a discrete multi-objective optimization problem in an efficient way. The neighborhood structure is updated with every new nondominated point. An advantage of this neighborhood structure is that once one local nadir point is known that has to be updated due to the insertion of a new point, one can easily navigate through the list of local nadir points to find all those that have to be updated in this iteration, too.

### 4.2.4 Proposed algorithm

We concluded that the following combination of algorithms might be promising.

As a baseline, we could use the algorithm from [8] to quickly identify the solution with the minimum HV contribution. As a result, we obtain bounds of the HV contribution of each solution in the population. After removing the solution with the minimum HV contribution and adding a new solution, we would have to update the information to quickly identify the next solution with minimum HV contribution. This would be done in two steps:

1. First, we would check whether there are any dominated solutions. If there is at least one dominated solution, we can remove one of them at random as all of their HV contributions are equal to zero, and thus minimal. If several dominated solutions are found, they can be removed iteratively without the need to recompute any HV contributions.
2. If there are no dominated solutions, we can use the algorithm in [5] to identify the solutions whose HV contribution may have changed. Only those solutions need to have their bounds re-set, while the other solutions may keep their bounds from the previous iteration. With this update, the algorithm from [8] can be used again to quickly identify the solution with minimal HV contribution.

We also noted that it is easy to adapt the algorithm in [8] to work with approximations, rather than running it until the solution with the guaranteed smallest HV contribution remains. One could stop the algorithm earlier, while some intervals still overlap, with the tolerated overlap controlling the approximation error.

The number of solutions that need to be updated in one iteration of Step 2 may vary from one iteration to another, but it can generally not be large in all iterations.

#### 4.2.5 Additional ideas

Additional ideas discussed at the working group include

- The paper by [2] proposes an efficient algorithm to identify the solution with minimal HV contribution and approximation guarantee. It would be worthwhile to explore whether it is possible to speed this up by transferring computational results from one iteration to the next.
- For algorithms that add or remove more than one solution, [4] discusses possibilities to speed up computations.
- For ray-based approximations: when increasing the budget to increase accuracy, is it better to use more rays or more points along each ray?
- Can we cleverly employ GPUs for MC estimation approaches for HV? Which way to cut? Should each sample be evaluated in a core containing the reference set, or each reference set member on a core compared to a set of samples?
- How accurate do we need the estimation to be for effective use in (expensive) optimisation algorithms – i.e., what budgets do we need to employ. Is this fixed throughout the run or should it vary? Can we reproduce the observations from [1] for noisy single-objective problems: Accuracy is important at the start, not really in the middle, very important at the end?
- What is our budget for approximation given known computation time for fitness evaluation, i.e., should we rather spend more time on accurate HV computations, or work with crude approximations and instead do more iterations?
- If we only add points to the Pareto front, or only remove points, then we can make use of the fact that HV contributions can only decrease or increase, respectively. But only adding points will mean that we need to have an unconstrained size of the Pareto front.

#### 4.2.6 Acknowledgements

This report was conducted as part of the “surrogates” subgroup of the Dagstuhl Seminar 23361 Multiobjective Optimization on a Budget, and the work benefited from wider discussions with Jonathan Fieldsend, Hisao Ishibuchi, Pascal Kerschke, Joshua Knowles, Boris Naujoks, Alma Rahat, Vanessa Volz, Hao Wang, Kaifeng Yang.

**References**

1 Branke, J. (2012). *Evolutionary optimization in dynamic environments*, volume 3. Springer Science & Business Media.

2 Bringmann, K. and Friedrich, T. (2012). Approximating the least hypervolume contributor: Np-hard in general, but fast in practice. *Theoretical Computer Science*, 425:104–116.

3 Chen, W., Ishibuchi, H., and Shang, K. (2020). Lazy greedy hypervolume subset selection from large candidate solution sets. In *2020 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8.

4 Chen, W., Ishibuchi, H., and Shang, K. (2021). Fast greedy subset selection from large candidate solution sets in evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 26(4):750–764.

5 Dächert, K., Klamroth, K., Lacour, R., and Vanderpooten, D. (2017). Efficient computation of the search region in multi-objective optimization. *European Journal of Operational Research*, 260(3):841–855.

6 Guerreiro, A. P., Fonseca, C. M., and Paquete, L. (2020). The hypervolume indicator: Problems and algorithms. *arXiv preprint arXiv:2005.00515*.

7 Jaszkiewicz, A., Susmaga, R., and Zielniewicz, P. (2020). Approximate hypervolume calculation with guaranteed or confidence bounds. In *Parallel Problem Solving from Nature–PPSN XVI: 16th International Conference, PPSN 2020, Leiden, The Netherlands, September 5-9, 2020, Proceedings, Part I 16*, pages 215–228. Springer.

8 Jaszkiewicz, A. and Zielniewicz, P. (2021). Quick extreme hypervolume contribution algorithm. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 412–420.

9 Jaszkiewicz, A. and Zielniewicz, P. (2022). Greedy decremental quick hypervolume subset selection algorithms. In *International Conference on Parallel Problem Solving from Nature*, pages 164–178. Springer.

10 Nan, Y., Shang, K., Ishibuchi, H., and He, L. (2022). An improved local search method for large-scale hypervolume subset selection. *IEEE Transactions on Evolutionary Computation.*

11 Shang, K., Ishibuchi, H., He, L., and Pang, L. M. (2021). A survey on the hypervolume indicator in evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 25(1):1–20.

12 Shang, K., Ishibuchi, H., and Ni, X. (2019). R2-based hypervolume contribution approximation. *IEEE Transactions on Evolutionary Computation*, 24(1):185–192.


## 4.3 Knowledge Extraction for Multiobjective Optimization on a Budget

*Tinkle Chugh (University of Exeter, GB), Karl Heinz Küfer (Fraunhofer ITWM – Kaiserslautern, DE), Kaisa Miettinen (University of Jyväskylä, FI), Markus Olhofer (HONDA Research Institute Europe GmbH – Offenbach, DE), Alma Rahat (Swansea University, GB)*



Real world optimization problems typically incorporate multiple objectives, and therefore generating and selecting a solution is not straightforward, and often leads to repeated interactions between decision makers and analysts before a final solution can be identified. Existing literature focus primarily on methodological contributions to elements of this overarching process, without elaborating the ecosystem within which interactive decisions are being made. Thus, there is a need to develop a concise and abstract frame of reference that allows discussions about these interactive knowledge exchanges, and in this report we present a skeleton structure to promote discussions in interactive decision-making in the context of multiobjective problem-solving on a budget.

### 4.3.1 Introduction

Multiobjective optimization refers to finding optimal solutions under the presence of multiple, often conflicting objective functions. Often, this is to be done within a predefined budget arising from resource limitations. The typical process consists of solving a multiobjective optimization problem specified by problem owners (POs) followed by decision makers (DMs) interacting with the solutions and studying the trade-offs between objective functions, and considering constraints (e.g., manufacturability of a mechanical apparatus), and ancillary information to identify solutions that meet their requirements. This is a rich field of study with many publications focusing on solving specific aspects of the whole process, see, for example, [1] for a survey on the various interactive multiobjective optimization methods, and [2] for a recent overview and survey of how to assess performance in this context. Nonetheless, there is a lack of an overall framework and taxonomy that allow discussion and contributions to be directed within a well-thought-out structure. As pointed out, e.g., in [3], the whole process with the problem formulation is rarely discussed in publications. In this report, we aim to propose an initial sketch for such a frame of reference that can be used as a tool by researchers to facilitate discussions and identify the scope of contributions within this structure. Next, we briefly discuss the proposed frame of reference.

### 4.3.2 Frame of Reference for Knowledge Exchange (FKE)

Multiobjective optimization and related decision-making is an iterative process, and there are many elements to it. Firstly, we have the DMs or POs, who are attempting to solve a decision-making problem with many objective functions, supported by an (or a team of) *analyst*(s), or in other words an expert in optimization and decision-making methodologies. Together they are the elements of the *human side* of the process. Naturally, the other component of this process is then the *computational side* consisting of a mathematical or computational model devised through a specification, a suite of solvers appropriates for addressing the multiobjective optimization problem (MOP), and a module for information and knowledge extraction (e.g., for visualising the trade-offs) with an interactive user interface.

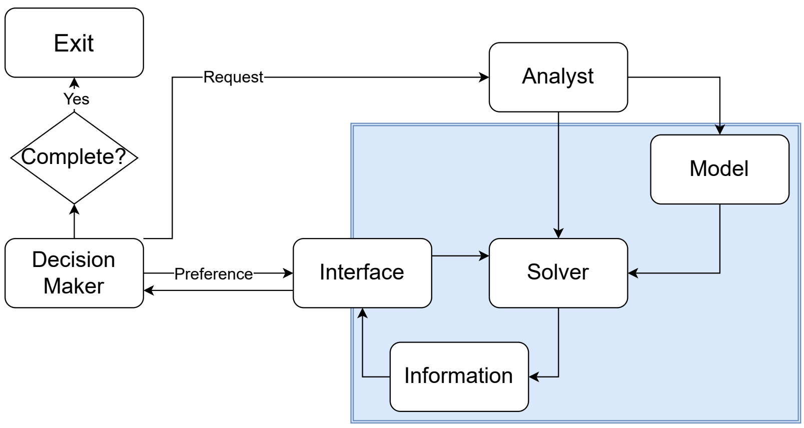



**Figure 6** The proposed frame of reference for knowledge exchange (FKE). The core concept captures the idea that the DM and the analyst work together on the *human side* in identifying the best solution, through cycles of requests and elicitations through model updates in the *computational world* (in blue shaded region), which is used by the solver to produce a set of solutions that the DM can interact with to reach final conclusions.

There are different forms of interactions between the elements. On the human side, the interactions are between the DM and the analyst. DMs use *requests*, to inform the analyst of what would be required of the model. Requests are related to the objective functions, constraints and decision variables. They can be reductionist (e.g., locate the best subset), expansive (e.g., include a new aspect), or a form of amendments (e.g., refinement or reformulations). Furthermore, the requests tend to be triggered by some recognition or need for change. The instigator in this case can be either the DM or the analyst. When the DM instigates, apart from the initial model specification, it may be because of new information becoming available, for example, a recognition of a new element of the problem that was not defined initially, a change in preference information relative to objective functions, new knowledge about the problem becoming available, or discrepancy with the real world requirements. On the other hand, an analyst may be able to instigate a change request after observing the behaviour of the DM. The latter may be automated through anomaly detection techniques.

At the boundary of the human-computer ecosystem, the analyst uses the requests to (re-)formulate model specifications and configure the appropriate solvers. On the other hand, the DMs interact with the computational side through a user interface where they can query the (estimated) Pareto front, and iteratively identify interesting (regions) of solutions. Moreover, they may instigate another run of the solver to investigate particular regions of interest through preference elicitation.

These core elements and interactions can be used to construct an *frame of reference for knowledge exchange* (FKE) that captures the solution process. An illustration of the IDF is shown in Figure 6. We now discuss the framework through an example.

#### 4.3.2.1 A first Example: Multiobjective Interactive Radiotherapy Assistant (MIRA)

A software tool was proposed in [4] known as a multiobjective Interactive Radiotherapy Assistant (MIRA). It is used for radiotherapy planning with multiobjective optimization through an interactive exploration of the solution space. In this tool, a radiologist identifies a target volume and an associate dosage. Typically, there will be millions of voxels (where each voxel is a collection of pixels in a volumetric image), with each representing an objective function. Therefore, only a subset of the objective functions is used to reflect an organ of interest. A dose distribution should immediately go down for a healthy organ to protect them, while other target parts follow a different distribution.

Considering FKE, the interactions between the DM and the analyst may occur as follows: a DM identifies the voxels and defines thresholds for exploration, and then the analyst configures the solver and the model to extract an approximate Pareto front. In addition, a DM can pick a point on the imageries to highlight a new voxel, and hence introduce a new objective function. On the other hand, the DM working with the interface can merely look at a subset of the currently generated front. Alternatively, they could also focus on a part of the front, and rerun the solver to improve the approximation.

#### 4.3.2.2 A Second Example: Multiobjective dynamic vehicle performance optimisation

A second example is given by the improvement of dynamic vehicle performance during automotive design processes [5]. The high-speed stability of a vehicle is crucial for comfortable drive during highway scenarios. It relates to stable, predictable and controllable vehicles and finally results in ride comfort and road holding capabilities. The vehicle behaviour is mainly influenced by the interaction between components in the suspension system, the steering subsystems and the tire characteristics, resulting in a high number of adjustment factors. Usually, these tests and tuning related to the comfort level of a vehicle are done late in the design stages with road registered prototype vehicles. However, with the introduction of virtual developments tools these test can be performed early in the design process and allow for much higher variety of tests. The subjective evaluation of a vehicle however requires the consideration of various scenarios during day to day driving which results in a high number of criteria for the optimisation. This results on the one hand in a high number of parameters describing the vehicle, on the other hand a high dimensional solution space in which most criteria are in a trade-off relationship. The selection of one single vehicle configuration requires therefore the identification of a preferred solution in the high dimensional search space. At the same time, it can become obvious, that evaluation scenarios needs to be exchanged or adapted during decision making and analysis processes, resulting in changes in the optimisation criteria or even in the underlying system structure, resulting in changes in the pasteurisation of the problem.

In the next section, we discuss briefly how different levels of budget relate to the frame of reference.

### 4.3.3 Context and Examples of Budgets

Different interactions may be associated with different types of budgets, but they are primarily due to resource limitations. The DM must identify the ultimate solution within any such budget. Below, we present a few examples of budgets in this context.

**Wall clock time.** The overall time before the DM must finalise their decision may be well-defined.

**Model related budget.** Many real world problems would require substantial time for each function evaluation if it requires numerical simulations. For example, a computational fluid dynamic simulation of a draft tube may take a thousand seconds [6]. So, the budget may be about a few hundred function evaluations, and thus impact the interactions and their nature.

**Solver budget.** The solver may itself be expensive. For example, entropy search for multiobjective optimization requires numerically approximating an acquisition function that can discriminate between solutions, but optimizing such acquisition functions to identify the next best solution may be exorbitant [7]. Thus, there are practical limitations on how much time we can spend before evaluating the next solution.

**Proprietery software.** Many professional simulation software are proprietary, and therefore, there may be limits to how many licences are available to a DM.

**Preparation budget.** Interactions between the DM and analysts may take an insignificant amount of time for discussions. In addition, for the analyst to evaluate and prepare models/solvers with DM guidance may take some time.

### 4.3.4 Conclusion

In this report, we briefly proposed and discussed a framework for iterative discovery of final solutions by a DM while interacting with multiobjective optimization methods supported by analysts. Future work involves expanding the framework and validating it with multiple real examples, with the possibility of incorporating multiple DMs working independently.

**References**

1 Bin Xin, Lu Chen, Jie Chen, Hisao Ishibuchi, Kaoru Hirota, and Bo Liu. Interactive multiobjective optimization: A review of the state-of-the-art. *IEEE Access*, 6:41256–41279, 2018.

2 Bekir Afsar, Kaisa Miettinen, and Francisco Ruiz. Assessing the performance of interactive multiobjective optimization methods: a survey. *ACM Computing Surveys*, 54(4):1–27, 2021.

3 K. Deb, P. Fleming, Y. Jin, K. Miettinen, and P. M. Reed. Key issues in real-world applications of many-objective optimisation and decision analysis. In D. Brockhoff, M. Emmerich, B. Naujoks, and R. Purshouse, editors, *Many-Criteria Optimization and Decision Analysis*, pages 29–57. Springer, 2023.

4 Christian Thieke, Karl-Heinz Küfer, Michael Monz, Alexander Scherrer, Fernando Alonso, Uwe Oelfke, Peter E Huber, Jürgen Debus, and Thomas Bortfeld. A new concept for interactive radiotherapy planning with multicriteria optimization: first clinical evaluation. *Radiotherapy and Oncology*, 85(2):292–298, 2007.

5 Martin Heiderich, Timo Friedrich, and Minh-Tri Nguyen. New approach for improvement of vehicle performance by using a simulation-based optimization and evaluation method. In *7th International Munich Chassis Symposium 2016: chassis.tech plus (Proceedings)*. Springer, LNCS, LNAI, LNBI, June 2016.

6 Steven Daniels, Alma Rahat, Gavin Tabor, Jonathan Fieldsend, and Richard Everson. Application of multi-objective bayesian shape optimisation to a sharp-heeled kaplan draft tube. *Optimization and Engineering*, pages 1–28.

7 Daniel Hernández-Lobato, Jose Hernandez-Lobato, Amar Shah, and Ryan Adams. Predictive entropy search for multi-objective bayesian optimization. In *International Conference on Machine Learning*, pages 1492–1501. PMLR, 2016.

## 4.4 Reducing Complexity in Multiobjective Optimization by Model Reduction

*Gabriele Eichfelder (TU Ilmenau, DE), Juliane Mueller (NREL – Golden, US), Enrico Rigoni (ESTECO SpA – Trieste, IT), Stefan Ruzika (RPTU – Kaiserslautern, DE), Michael Stiglmayr (Universität Wuppertal, DE)*



### 4.4.1 Introduction

Optimization problems have often to be solved on a limited budget, e.g. due to time restriction or restrictions on the number of function evaluations. Multiobjective optimization problems are in general more costly to solve compared to single-objective optimization problems. Specifically, the number of objective functions and the type of objective functions directly influence the effort needed to solve the problem. Hence, in case of budget constraints for solving the problem, an important first step is to reduce the complexity of the multiobjective model as far as possible. In our working group, we have discussed and analyzed different ways to simplify a multiobjective optimization problem using reduction approaches.

We considered both reduction of structural complexity (e.g. by low-order polynomial approximation of objective functions) and reduction of problem size, in particular with respect to objective functions (e.g. by scalarization and aggregation of different objective functions). We refer to Figure 7 for an overview of different possible reduction scopes.

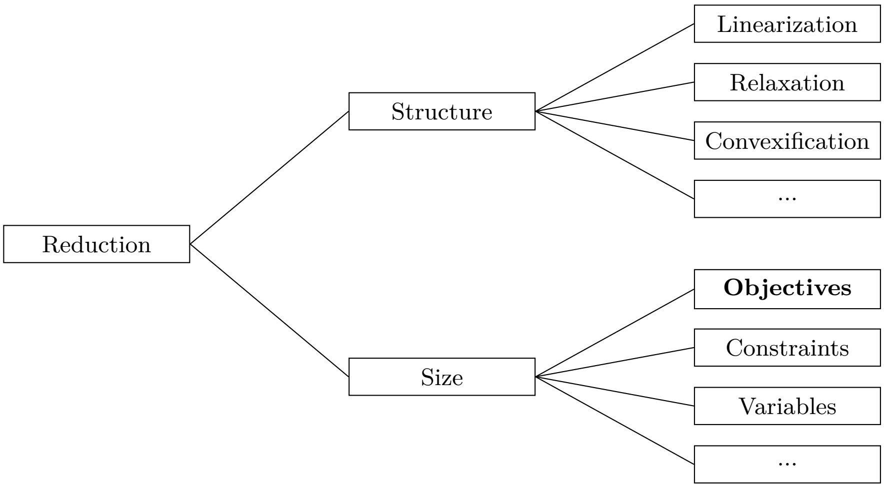



**Figure 7** Overview of different possible reduction scopes. In this report, we focus on the reduction of the number of objectives by means of scalarization and aggregation of different objective functions.

As mentioned above, reducing the complexity of a multiobjective problem is relevant in the context of optimization on a budget, with the idea of making better use of the limited number of evaluations available. The price of the reduction in general as well as measures for the quality of some reduction in particular are other interesting topics that deserve to be explored.

In this report, we present the main results achieved during the seminar:

- Reducing the number of objective functions by scalarizing some of the convex objectives – refer to Section 4.4.2
- Sufficient conditions for reducing the number of objectives locally using gradients – refer to Section 4.4.3

For completeness, we list other interesting discussion topics that we did not have time to delve into due to the limited time resources available:

- Reducing the complexity of the problem by (local) low-order polynomial approximation of objective functions (linearization or quadratic approximation)
- Best balance of high- and low-fidelity models during optimization
- Possible use of surrogate models to check local convexity

### 4.4.2 Reducing the Number of Objective Functions by Convex Combination

We aimed at the examination of the relation between the efficient set, i.e. the set of efficient solutions of a tri-objective problem and a family of bi-objective problems, obtained by combining two of the functions using weighted sums and by varying the weights. To be more specific, we examine the problem

$$\min_{x \in S} \begin{pmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{pmatrix} \tag{MOP}$$

with $f_i\colon \mathbb{R}^n \to \mathbb{R}$, $i \in [3] := \{1, 2, 3\}$ continuous functions and a feasible set $S \subseteq \mathbb{R}^n$. The set of efficient solutions of (MOP) is denoted by $\mathcal{E}$ and the set of weakly efficient solutions by $\mathcal{E}_w$.

With this problem, we associate the family of problems

$$\min_{x\in S} \begin{pmatrix} f_k(x) \\ wf_i(x)+(1-w)f_j(x) \end{pmatrix} \tag{BOP($k,w$)}$$

where $k \in [3]$ and $w \in [0,1]$. The set $\{i,j\} := [3]\setminus\{k\}$ refers to the remaining indices, i.e. the indices other than $k$. The set of efficient solutions of (BOP($k,w$)) is denoted by $\mathcal{E}(k,w)$ and the set of weakly efficient solutions by $\mathcal{E}_w(k,w)$. In the literature, the effects of adding or deleting an objective function have been studied, see [2]. Moreover, for a fixed choice of $k$ and $w$, (BOP($k,w$)) and its relation to (MOP) has been addressed among others in [1]. Here, we examine the family of problems (BOP($k,w$)).

The following result relates the set of optimal solutions of the problems mention above to each other:

▶ **Theorem 1.** *Let the set $f(S)+\mathbb{R}^3_+$ be convex (which, for instance, holds true if the functions $f_i\colon \mathbb{R}^n \to \mathbb{R}$, $i \in [3]$ and the set $S \subseteq \mathbb{R}^n$ are both convex). Then, for any $k \in [3]$, it holds*

$$\bigcup_{w\in(0,1)} \mathcal{E}(k,w) \subseteq \mathcal{E} \subseteq \bigcup_{w\in[0,1]} \mathcal{E}_\omega(k,w).$$

**Proof.** We start with the first inclusion. W.l.o.g. let $k=1$ and then $i=2$, $j=3$. Let $w \in (0,1)$ and $\bar{x} \in \mathcal{E}(1,w)$. Assume $\bar{x} \notin \mathcal{E}$. Then, there exists $\tilde{x} \in S$ with $f_1(\tilde{x}) \le f_1(\bar{x})$, $f_2(\tilde{x}) \le f_2(\bar{x})$, and $f_3(\tilde{x}) \le f_3(\bar{x})$, with strict inequality for at lest one of the inequalities. Thus, it holds $wf_2(\tilde{x})+(1-w)f_3(\tilde{x}) \le wf_2(\bar{x})+(1-w)f_3(\bar{x})$.

If $f_1(\tilde{x}) < f_1(\bar{x})$, then this is a contradiction to $\bar{x} \in \mathcal{E}(1,w)$.

Otherwise, $f_2(\tilde{x}) < f_2(\bar{x})$ or $f_3(\tilde{x}) < f_3(\bar{x})$ holds true. We have $wf_2(\tilde{x})+(1-w)f_3(\tilde{x}) < wf_2(\bar{x})+(1-w)f_3(\bar{x})$ in contradiction to $\bar{x} \in \mathcal{E}(1,w)$.

For the second inclusion, let $\bar{x} \in \mathcal{E}$. Then, due to convexity of $f(S)+\mathbb{R}^3_+$, there exists a vector $v \in \mathbb{R}^3_+ \setminus \{0\}$ such that $\bar{x}$ is a minimal solution of

$$\min_{x\in S} v_1 f_1(x) + v_2 f_2(x) + v_3 f_3(x).$$

Set $\alpha := v_2 + v_3 \ge 0$. If $\alpha = 0$, then $v_2 = v_3 = 0$, $v_1 > 0$ and $\bar{x}$ is a minimal solution of

$$\min_{x\in S} v_1 f_1(x)$$

and, thus, $\bar{x} \in \mathcal{E}_\omega(k,w)$. Otherwise, $\alpha > 0$ and $\bar{x}$ is also a minimal solution of

$$\min_{x\in S} \frac{v_1}{\alpha} f_1(x) + \frac{v_2}{\alpha} f_2(x) + \frac{v_3}{\alpha} f_3(x)$$

and for $w := \frac{v_2}{\alpha} \in [0,1]$, the point $\bar{x}$ is a minimal solution of a weighted sum ob the objectives $f_1(x)$ and $wf_2(x)+(1-w)f_3(x)$ with the two weights $v_1/\alpha \ge 0$ and 1. Thus, $\bar{x} \in \mathcal{E}_\omega(k,w)$. ◀

However, the biobjective problems (BOP($k,w$)) are in general not capable of covering the full complexity of the three-objective problem (MOP) as the following counterexample within the next proposition shows.

▶ Proposition 1. Efficient solutions of (MOP) are not necessarily efficient for one of the associated biobjective problems (BOP($k,w$)), i.e. there may exist an efficient solution $x \in \mathcal{E}$ such that $x \notin \mathcal{E}(k,w)$ for all $k \in [3]$ and $w \in [0,1]$.

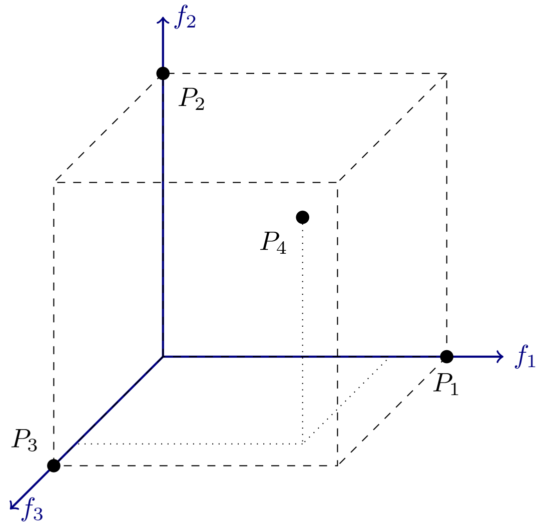



**Figure 8** For all biobjective weighted sum problems of the form (BOP($k,w$)), $k \in \{1,2,3\}$, the point $P_4$ is dominated.

**Proof.** For the following counterexample, we consider the points in outcome space. Let four feasible points in outcome space be given by $P_1 = (5,0,0)^\top$, $P_2 = (0,5,0)^\top$, $P_3 = (0,0,5)^\top$, $P_4 = (4,4,4)^\top$, see Figure 8. Then, $P_4$ is dominated for all problems of the form (BOP($k,w$)) (i.e., for all $k \in \{1,2,3\}$):

For $k = 1$, the four points are mapped to $\binom{5}{0}$, $\binom{0}{5\,w}$, $\binom{0}{5\,(1-w)}$, $\binom{4}{4}$. Then, $\binom{4}{4}$ is dominated for all $w \in [0,1]$ by $\binom{0}{5\,w}$ or by $\binom{0}{5\,(1-w)}$. The cases, $k = 2, 3$ yield the same result due to symmetry. ◀

The proof of Proposition 1 shows that unsupported efficient solutions may not be obtained as optimal solutions of one of the associated biobjective subproblems (BOP($k,w$)). Thus, additional convexity assumptions seem to be necessary. However, as the following proposition shows, it is not sufficient if only two of the three objectives are convex.

▶ Proposition 2 ([3]). Consider the problem (MOP). Moreover, we assume that $S$ is a convex set, $f_2$ and $f_3$ are convex functions. Then, there may exist efficient solutions $x \in \mathcal{E}$ for (MOP) which can not be obtained as efficient solutions of (BOP(1,$w$)) with $w \in [0,1]$, i.e., $x \notin \bigcup_{w\in[0,1]} \mathcal{E}(k,w)$.

**Proof.** Consider the tri-objective problem (MOP) with $n = 2$, $f_1(x) = (1-(x_1-1)^2-x_2)^2$, $f_2(x) = x_1+\varepsilon\,x_2$, $f_3(x) = 2-x_1+\varepsilon\,x_2$ for some $\varepsilon > 0$. Moreover, let $S = \{(x_1,x_2)^\top : x_1,x_2 \geq 0\}$, i.e. we study

$$\min_{x_1,x_2\geq 0} \begin{pmatrix} (1-(x_1-1)^2-x_2)^2 \\ x_1+\varepsilon\,x_2 \\ 2-x_1+\varepsilon\,x_2 \end{pmatrix} \tag{7}$$

Then, $\bar{x} = (1,1)^\top$, $\hat{x} = (0,0)^\top$ and $x' = (2,0)^\top$ are efficient solutions, since the non-negative objective function $f_1$ equals zero for all of them (i.e., $f_1(\bar{x}) = f_1(\hat{x}) = f_1(x') = 0$). The corresponding vectors in the outcome space are

$$f(\bar{x}) = \begin{pmatrix} 0 \\ 1+\varepsilon \\ 1+\varepsilon \end{pmatrix} \qquad f(\hat{x}) = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} \qquad f(x') = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$$

For $\bar{x}, \hat{x}, x'$, we determine the corresponding outcome vectors for the bi-objective optimization problem (BOP(1,$w$))

$$\begin{pmatrix} f_1(\bar{x}) \\ w\, f_2(\bar{x}) + (1-w)\, f_3(\bar{x}) \end{pmatrix} = \begin{pmatrix} 0 \\ 1+\varepsilon \end{pmatrix}$$
$$\begin{pmatrix} f_1(\hat{x}) \\ w\, f_2(\hat{x}) + (1-w)\, f_3(\hat{x}) \end{pmatrix} = \begin{pmatrix} 0 \\ 2\,(1-w) \end{pmatrix}$$
$$\begin{pmatrix} f_1(x') \\ w\, f_2(x') + (1-w)\, f_3(x') \end{pmatrix} = \begin{pmatrix} 0 \\ 2\,w \end{pmatrix}$$

One can easily verify, that $\bar{x}$ is a dominated solution of (BOP(1,$w$)) for all values $w \in [0,1]$, since $\bar{x}$ is a dominated by $\hat{x}$ if $w > \frac{1}{2}(1-\varepsilon)$ or by $x'$ if $w < \frac{1}{2}(1+\varepsilon)$. ◀

Note that similar results can be easily shown for $k = 2, 3$, since the weighted sum in the second objective of (BOP($k, w$)) involves a potentially non-convex function.

### 4.4.3 Descent Algorithms

Let us consider an algorithm for solving the multiobjective optimization problem (MOP) and assume that this algorithm relies on iteratively computing descent directions of the individual objective functions in a local optimization procedure. Then, if the gradient of one individual objective function is locally a convex combination of the others, this objective function does not have to be considered in the optimization process. More precisely, the following statement holds true.

We use the definition that a direction $d$ is a descent direction for a continuously differentiable function $g\colon \mathbb{R}^n \to \mathbb{R}$ in $\bar{x}$ if $\nabla g(\bar{x})^\top d < 0$.

▶ **Lemma 2.** *Consider* (MOP) *as above with continuously differentiable objective functions and $S = \mathbb{R}^n$. Suppose there is $\bar{x} \in S$ and $\mu \in [0,1]$ such that $\nabla f_3(\bar{x}) = \mu \nabla f_1(\bar{x}) + (1-\mu) f_2(\bar{x})$. Then, any descent direction $d$ for $f_1$ and $f_2$ in $\bar{x}$ is also a descent direction for $f_3$ in $\bar{x}$.*

**Proof.** Using that $d$ is a descent direction for $f_1$ and $f_2$ we immediately get

$$\nabla f_3(\bar{x})^\top d = \mu \nabla f_1(\bar{x})^\top d + (1-\mu) f_2(\bar{x})^\top d < 0.$$ ◀

**References**

1 Dempe, S., Eichfelder, G., and Fliege, J.: *On the effects of combining objectives in multi-objective optimization.* Mathematical Methods of Operations Research, 82(1), 1-18, 2015.
2 Fliege, J.: *The effects of adding objectives to an optimisation problem on the solution set.* Operations Research Letters, 36(6), 782-790, 2017.
3 Klamroth, K.: Personal Communication, 2023.

## 4.5 Rank-based Surrogates for Multiobjective Optimization

*Ekhine Irurozki (Telecom Paris, FR), Dimo Brockhoff (INRIA Saclay – Palaiseau, FR), Tea Tusar (Jozef Stefan Institute – Ljubljana, SI)*



### 4.5.1 Introduction

Some decisions within current (evolutionary) multiobjective optimization algorithms are solely based on rankings among solutions. The environmental selection step is one classical example. In the case where these decisions are costly, for example because the objective functions playing a role in the decision are expensive to evaluate, surrogate models for this decision process can be learned to replace the costly exact decision process by a cheaper, hopefully adequate enough process. This is in contrast with most existing surrogate-assisted (multiobjective) algorithms in which the objective functions are modelled by surrogates and the decision process is run on the surrogate-evaluated solutions.

In the subgroup on rank-based surrogates, we explored the possibilities of how rank-based surrogates can be used when rank-based decisions have to be made within an algorithm. As a first simple example, we propose to use a support vector machine to predict the environmental selection decisions in the NSGA-II algorithm [2]. In the following, we discuss the main ideas behind this rank-surrogate based NSGA-II and provide its pseudocode. Implementation, testing, and numerical benchmarking of the proposed algorithm remains a task for future work.

### 4.5.2 The Proposed Algorithm

Following the ideas of the lq-CMA-ES algorithm [3] for single-objective surrogate-based optimization, we discuss a very basic multiobjective algorithm with a rank-based surrogate. The main working principle of the algorithm is to learn a rank-based surrogate model for the decision which solutions to keep and which to abandon at each step (environmental selection). The pseudocode in Algorithm 2 uses as baseline algorithm the well-known NSGA-II [2] and a Support Vector Machine (SVM, [1]) as the surrogate model but other multiobjective algorithms and surrogate models could be used instead as well.

In order to save as many expensive function evaluations as possible, we evaluate, in each iteration, only a small proportion $p_\lambda$ of the $\lambda$ newly sampled solutions. Based on the last $\lambda$ evaluated solutions in an archive $\mathcal{A}$ and the corresponding environmental selection decision on them, a surrogate model is learned to predict, for each solution $x_i$ of a new set of $\lambda$ solutions $x_1, \ldots, x_\lambda$ whether the solution $x_i$ is to be kept (value 1) or not (value 0) for the next iteration. If the new model is predicting the environmental selection decisions well enough compared to the previous model (in terms of the Kendall-tau rank correlation coefficient, [4], line 15), we save the remaining function evaluations of the iteration and continue with the original algorithm, here NSGA-II. Only if the predictions between the old and the new surrogate model differ too much, i.e., if the Kendall rank correlation coefficient is smaller than a given threshold $T_\tau$, the next $\lceil p_\lambda \lambda \rceil$ solutions of the current iteration are evaluated successively on the true objective functions, and a new model is learned until the rank correlation coefficient between the old and the new model is larger than the target $T_\tau$ (or until all $\lambda$ solutions in the iteration are evaluated).

Note that in the pseudocode of Algorithm 2, the classifier $C_t$ returns $\mu$ solutions out of a set of $\mu + \lambda$ solutions (i.e., the ones that are selected for survival). To achieve this with SVMs, we return the $\mu$ solutions with the largest predicted value among all $\mu + \lambda$ solutions.

**Algorithm 2:** NSGA-II with environmental selection replaced with SVM prediction.

Inputs:
- $\mu$: population size of the multiobjective algorithm
- $\lambda$: number of offspring per iteration
- $p_\lambda$: percentage of offspring evaluated together (default: 10%)
- $T_\tau$: threshold for Kendall-$\tau$ comparison (default: 85%)
- $n_0$: size of initial sampling/DOE

$t \leftarrow 0$, init population $X_0$ with $n_0$ points (via DOE or uniformly at random)<br>
$y_i \leftarrow f(x_i)$ for all $x_i \in X_0$, store results $(x_i, y_i)$ in archive $\mathcal{A}$<br>
$X_1 \leftarrow$ mutation_crossover_NSGA–II($X_0$) ▷ with random mating selection for now<br>
▷ and $\mu := |X_t| = |X_{t+1}| =: \lambda$<br>
$y_i \leftarrow f(x_i)$ for all $x_i \in X_1$, add results $(x_i, y_i)$ to archive $\mathcal{A}$<br>
$X_1' \leftarrow$ environmentalselection($\mathcal{A}$)<br>
Train classifier $C_1$ based on $\mathcal{A}$ and $X_1'$<br>
$X_1 \leftarrow X_1'$, $t \leftarrow t + 1$,<br>
while not happy:<br>
    $X_{t+1} \leftarrow$ mutation_crossover_NSGA-II($X_t$) ▷ as above<br>
    for $j$ in $\{1, \ldots, \lfloor 1/p_\lambda \rfloor\}$:<br>
        for all $i \in \{(j-1) \cdot \lceil p_\lambda \lambda \rceil + 1, \ldots, j \cdot \lceil p_\lambda \lambda \rceil\}$ do:<br>
            $y_i \leftarrow f(x_i)$ and replace oldest entry in archive $\mathcal{A}$ with $(x_i, y_i)$<br>
        Train classifier $C_{t+1}$ based on $\mathcal{A}$ and environmentalselection($\mathcal{A}$)<br>
        If Kendall-$\tau(C_t(X_t \cup X_{t+1}), C_{t+1}(X_t \cup X_{t+1})) > T_\tau$ ▷ Classifier good enough<br>
                or if all solutions are already evaluated:<br>
            $X_{t+1} \leftarrow C_{t+1}(X_t \cup X_{t+1})$, $t \leftarrow t + 1$, break

### 4.5.3 Final Notes

Rank-based decisions in randomized algorithms such as the environmental selection of evolutionary algorithms might profit from rank-based surrogates that are only trained on and can only provide rankings. If the algorithm is invariant to monotonous transformations of the objective functions, for example, it will keep this property with a rank-based surrogate. Our proposal of using a support vector machine to predict the environmental selection within NSGA-II is the first step towards this goal.

It remains to be shown that such a simple rank-based surrogate actually works in practice. Potentially, we need to feed more information to the surrogate model than just the decision of whether a solution is kept or not. Pairwise rankings between solutions or a total ranking on the input solution set can be imagined here.

#### References


1 Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Computational Learning Theory (COLT)*, COLT '92, pages 144–152. ACM, 1992.

2 Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.

3 Nikolaus Hansen. A Global Surrogate Assisted CMA-ES. In *GECCO 2019 – The Genetic and Evolutionary Computation Conference*, pages 664–672, Prague, Czech Republic, July 2019. ACM.

4 M. G. Kendall. A New Measure of Rank Correlation. *Biometrika*, 30(1–2):81–93, 06 1938.

## 4.6 Computing R2-Optimal Representations of the Pareto Front on a Budget

*Kathrin Klamroth (Universität Wuppertal, DE), Fritz Bökler (Universität Osnabrück, DE), Susan R. Hunter (Purdue University, US), Andrzej Jaszkiewicz (Poznan University of Technology, PL), Arnaud Liefooghe (University of Lille, FR), Luís Paquete (University of Coimbra, PT), Serpil Sayin (Koc University – Istanbul, TR), Britta Schulze (Universität Wuppertal, DE), Daniel Vanderpooten (University Paris-Dauphine, FR)*



### 4.6.1 Introduction

Decision-making problems under conflicting objectives can be modeled and solved using multiobjective optimization. In general, such optimization problems with two or more objectives do not lead to a single optimal solution; instead, they lead to a whole set of so-called efficient solutions and corresponding nondominated points in the objective space. The set of all corresponding nondominated points is often called the Pareto front. Since each point cannot be improved in one or more objectives without worsening in at least one of the other objectives, all of these points are incomparable with respect to the optimization problem itself. Thus, usually in practice a decision maker must decide which solution is the best; such a decision should be made within the application context.

Decision-makers may bring inherent limitations to the optimization process, for example, in their availability and responsiveness. Furthermore, the set of nondominated points of a multiobjective optimization problem may become very large and, therefore, difficult for a decision maker to assort. In the following work, we present ideas that assist the decision-making process by creating a concise representation of the Pareto front that contains only a pre-determined number of nondominated points. That is, we assume the decision maker knows in advance a "budget" of how many nondominated points to consider when making a decision. The task is then to find a "good" representation of the Pareto front subject to the budget, so that the representation includes at most the given number of points. We consider a representation to be "good" if it is optimal with respect to some performance indicator of interest; further, the representation should not require advance computation of the complete Pareto front (or advance computation of a representation or approximation containing far more than the given number of points). The ability to compute such representations may improve decision-making by presenting a few options rather than all options, and may provide improved storage performance when multiple Pareto fronts must be computed as sub-problems of some larger multiobjective optimization (e.g., stochastic multistage) problem.

The remainder of this report is organized as follows. In the rest of the introduction, we provide a short overview on basic definitions and notation of multiobjective optimization and representation (Section 4.6.1.1), state the representation problem on a budget in a formal way (Section 4.6.1.2), and present previous work and existing literature (Section 4.6.1.3). In Section 4.6.2, we concentrate on the R2-indicator as quality measure for the representation. We provide an overview of the R2-indicator (Section 4.6.2.1), present a problem definition where the whole representation is computed "all-at-once" (Section 4.6.2.2), and present a mixed-integer reformulation of the problem (Section 4.6.2.3). The special case of multiobjective (mixed-integer) linear problems is discussed in Section 4.6.2.4. Computational results are included in Section 4.6.3, and Section 4.6.4 contains concluding remarks and future research.

#### 4.6.1.1 Basic Definitions, Assumptions, and Notation

We consider the context of solving a multiobjective optimization problem

$$\begin{aligned} \text{maximize} \quad & f(x) = (f_1(x), \ldots, f_q(x)) \\ \text{s.t.} \quad & x \in \mathcal{X} \subset \mathbb{R}^n, \end{aligned} \tag{8}$$

where at least two of the $q$ objective functions are conflicting and the feasible set $\mathcal{X}$ is nonempty. The solution to (8) is the efficient set, denoted by $\mathcal{X}_{\text{eff}}$, that consists of all the feasible points $x \in \mathcal{X}$ that can not be improved in all or some objectives without deterioration in at least one objective. More formally, the set of all *efficient* or *Pareto optimal* points is

$$\mathcal{X}_{\text{eff}} := \{x \in \mathcal{X} : \nexists x' \in \mathcal{X} \text{ such that } f(x') \geqslant f(x)\},$$

where whenever we compare two vectors $y, y' \in \mathbb{R}^q$, we write $y' \geqslant y$ to indicate that $y'_j \geqslant y_j$ for all $j = 1, \ldots, q$ and $y' \neq y$. (We use $y' \geqq y$ when equality is allowed.) The image $f(x)$ of an efficient point $x \in \mathcal{X}_{\text{eff}}$ is called *nondominated*, and the set of all nondominated points in the objective space is referred to as the *nondominated set*, *Pareto set*, or *Pareto front*,

$$\mathcal{Y}_{\text{Par}} = f(\mathcal{X}_{\text{eff}}) := \{f(x) : x \in \mathcal{X}_{\text{eff}}\}.$$

We refer the reader to the books [8, 15] for a thorough introduction to the field of multiobjective optimization.

Throughout the remainder of this work, we assume that the Pareto front $\mathcal{Y}_{\text{Par}}$ is nonempty and bounded. Under this assumption, without loss of generality, we further assume that all nondominated points are nonnegative in all components. That is,

$$\mathcal{Y}_{\text{Par}} \subset \mathbb{R}^q_{\geqq} := \{y \in \mathbb{R}^q : y \geqq 0\}; \tag{9}$$

in what follows, let $\mathbb{R}^q_{\geqslant}$ and $\mathbb{R}^q_{>}$ be defined accordingly.

We remark here that we consider a *representation* of the Pareto front, rather than an approximation. In [18], a discrete representation of the Pareto front is defined as a finite number of points selected from the Pareto front. In this sense, a representation of the Pareto front may differ from an approximation to the Pareto front, since an approximation may contain points that are dominated or otherwise do not belong to the Pareto front.

The notion of $\epsilon$-efficiency and the related concept of $\epsilon$-approximation [17] are well-known performance measures for approximations (see, for example, [9]); performance indicators to evaluate how close the representation is to the true Pareto front are also proposed. For general overviews of performance indicators, see [1, 13]. While some of these indicators require the Pareto front as an input, the efficient set and the Pareto front are often unknown a priori, so any methods developed to obtain representations cannot necessarily exploit such information while building the representation.

#### 4.6.1.2 Problem Statement

In the context of the multiobjective optimization problem (8), we consider the following problem statement.

▶ Problem 1. [Representation on a Budget] Suppose we are *given* an instance of problem (8) with unknown efficient set and unknown Pareto front, a performance indicator (for example, hypervolume, R2, coverage error, etc.), and a fixed budget $T$ for the number of representative points. Then, our problem is to *find* a representation of the Pareto front $\mathcal{R} \subseteq \mathcal{Y}_{\text{Par}}$ with cardinality $|\mathcal{R}| \leq T$ for which no other representation $\mathcal{R}'$ achieves a better performance with respect to the given performance indicator.

We remark that the budget of points $T$ applies only to the number of the points in the representation and does not otherwise limit the number of evaluations for the function $f$. In other words, in our case, the budget applies to the cognitive burden of the decision maker rather than to the computational resources.

Furthermore, our aim is to compute the whole representation at once ("all-at-once") such that we compute the representation "offline" and receive a guarantee that the entire representation $\mathcal{R}$ optimizes the performance indicator. Another approach would be to follow an iterative procedure, adding one point after the other to the representation ("one-by-one"); we use this term to encompass any sequential procedure, even if it adds more than one point at a time in an "online" fashion. Depending on the procedure and the underlying performance indicator, it may also be possible to provide a guarantee in the one-by-one case.

#### 4.6.1.3 Previous Work and Related Literature

We categorize previous work and related literature by whether the representation is determined all-at-once or one-by-one, as defined in Section 4.6.1.2. In addition to all-at-once and one-by-one, some procedures also require knowledge of the Pareto front in advance, or a finer representation before the representation of desired cardinality is constructed. Other approaches, not based on the optimization of a specific performance indicator, use binary relations relaxing the dominance relation, so as to control the quality of the representation in terms of coverage of the points that are not part of the representation set while ensuring a diversity among the points belonging to the representation set. This idea was implemented in [2] through the concept of $\varepsilon$-kernel. In this work, we take an all-at-once approach and we do not assume a known Pareto front or pre-construct a finer representation.

**All-at-Once.** The problem of finding an optimal representation of the Pareto front with respect to some performance indicator all-at-once, that is, so that the selected point set is (globally) optimal with respect to the chosen performance indicator, is also referred to as the *subset selection problem*. In this case, it is usually assumed that the complete Pareto front $\mathcal{Y}_{\text{Par}}$ is known beforehand. We refer to [10] for a recent review on this topic. In this context, the hypervolume indicator is frequently used to assess the representation quality. A closed formulation for the problem of finding $T$ representative points all-at-once without knowledge of the Pareto front with respect to the hypervolume indicator is given in [21] regarding the bi-objective fixed cardinality knapsack problem.

In the study of [25], the problem of finding a representative subset of nondominated points with respect to different combinations of quality measures is modeled as a multi-objective problem itself. Based on the knowledge of the nondominated set, these representation problems can be formulated as facility location problems with a special structure in the locations which makes the bi-objective problems solvable in polynomial-time.

**One-by-One.** For multiobjective linear programs, [19] provides an approach that aims to find discrete representations of the Pareto front. The approach suggests adding one new element to the representation of the Pareto front in each iteration with a control over the coverage error and it can terminate when $T$ Pareto points are obtained or the *coverage error* of the representation meets a target level. To implement this approach, the efficient faces are assumed to be known and a Mixed Integer Linear Programming (MILP) formulation per face is solved to obtain the representation. Two related studies that seek diverse subsets of the Pareto front are given by [24] and [14]. Both approaches are iterative in nature and build a representation by adding one nondominated point at a time. In [24], the goal is to locate a next point that has the largest Chebyshev distance to the region dominated by earlier points. The algorithm may terminate when this distance becomes acceptable or when $T$ points are obtained. Similarly, [14] proposes an approach that finds the next nondominated point that is at maximal Chebyshev distance from the existing points. Both approaches rely on a single MILP formulation to achieve their respective goals. As the number of points in the representation grows, the MILP formulation grows as well and may become computationally expensive. In [5], a variant with a better computational profile is presented. Moreover, [5] develops an approach that finds $T$ nondominated points simultaneously. This approach suggests building an approximation of the nondominated set first and obtaining its representation in a second step. Then nondominated points that are close to the approximate ones are identified.

In [20], an algorithm that is based on the Chebychev scalarization is introduced for biobjective discrete optimization problems. The algorithm stops when a prespecified coverage error is met by the representation that it builds. In [11], an algorithm that is based on the $\varepsilon$-constraint scalarization is given for biobjective discrete optimization problems. The study introduces the *representation error* as a performance indicator and uses it as a stopping condition.

A work that aims for an optimal representation of fixed cardinality with respect to the hypervolume indicator, that measures the dominated hypervolume of the point set with respect to some given reference point, is given by [16]. A hypervolume scalarization is used to define a sequence of, for bi-objective problems quadratic, optimization problems that provide one nondominated point at a time without assuming prior knowledge of the set $\mathcal{Y}_{\mathrm{Par}}$. The representation then yields a $(1-\frac{1}{e})$-approximation to the optimal representation in terms of the hypervolume indicator.

### 4.6.2 R2-Representation on a Budget

The selection of an appropriate performance indicator is crucial when computing representations of the Pareto front. In this context, the R2-indicator has recently received attention since it is relatively easy to evaluate, and since it yields representations that also perform well with respect to the hypervolume indicator [7]. We provide a formal definition below and refer to [12] and [4] for a more detailed introduction to the concept.

#### 4.6.2.1 Introduction to the R2-Indicator

The R2-indicator relies on weighted Chebyshev functions that can be defined as follows (note that problem (8) is a maximization problem):

▶ **Definition 1** (Weighted Chebyshev function). The *weighted Chebyshev (scalarizing) function* for a feasible solution $x \in \mathcal{X}$ of problem (8) is defined in the following way:

$$s_\infty(x,u,w) = \max_{j=1,\dots,q} w_j(u_j - f_j(x^t)) \tag{10}$$

where $u$ is a reference point such that $u \geqq f(x)$, and $w \geqslant 0$ is a nonzero and nonnegative weight vector. The map $s_\infty$ may be interpreted as a model of the decision maker's preferences – maximum weighted deviation from the reference point. The direction

$$\left(-\frac{1}{w_1},\dots,-\frac{1}{w_q}\right) \tag{11}$$

is called the *diagonal direction* of $s_\infty$ w.r.t. the reference point $u$.

A common choice for the reference point $u$ would be the ideal point or a utopia point of (8). Note that the value of the weighted Chebyshev function is the maximum weighted deviation from the reference point. Each weighted Chebyshev scalarizing function has at least one global optimum (minimum) belonging to the set $\mathcal{X}_{\text{eff}}$ of efficient solutions. For each efficient solution $x \in \mathcal{X}_{\text{eff}}$ there exists a weighted Chebyshev scalarizing function such that $s$ is a global optimum (minimum) of this function (Ch. 14.8 in [23]).

▶ **Definition 2** (R2-indicator). Given a set $\mathcal{S}$ of mutually non-dominated points in the objective space $\mathbb{R}^q$ and a reference point $u$ such that $u \geqq f(x^t)$ for all $f(x^t) \in \mathcal{S}$, the R2-indicator of $\mathcal{S}$ is the expected value over a set of normalized weight vectors $\Psi$ of the minimum value of the weighted Chebyshev functions achieved in the set $\mathcal{S}$ [4, 12]:

$$R2(\mathcal{S}, u) = \mathrm{E}_{w\in\Psi}[s^*_\infty(\mathcal{S}, u, w)] = \int_{w\in\Psi} s^*_\infty(\mathcal{S}, u, w)p(w)\, dw \tag{12}$$

where $p$ is a probability distribution function on $\Psi$ and

$$s^*_\infty(\mathcal{S}, u, w) = \min_{f(x^t)\in\mathcal{S}} s_\infty(x^t, u, w) = \min_{f(x^t)\in\mathcal{S}} \max_{j=1,\ldots,q} w_j(u_j - f_j(x^t)). \tag{13}$$

Since $s_\infty$ is a model of decision maker's preferences, the R2-indicator may be interpreted as the expected best utility achieved over $\mathcal{S}$ for all possible preferences. This quality indicator has been also independently proposed in [3] under the name *Integrated Preference Functional*. In the following we assume that $p$ is a uniform probability distribution function on $\Psi$.

In [7, 22], it has been proved that the R2-indicator is equivalent to the hypervolume indicator if $\Psi$ is a set of weight vectors corresponding to the set of diagonal directions uniformly distributed in polar coordinates and normalized with the $\ell_2$ norm, and the Chebyshev function is changed to its inverse form.

The expected value needed to calculate the R2-indicator may be estimated by an average over a finite sample of uniformly distributed weight vectors $\mathcal{W} \subseteq \mathbb{R}^q_{\geqslant}$, with

$$\mathcal{W} = \{w^1, \ldots, w^K\}$$

for $K$ large. Then we can calculate

$$R2(\mathcal{S}, u, \mathcal{W}) = \mathrm{E}_{w\in\mathcal{W}}[s^*_\infty(\mathcal{S}, u, w)] = \frac{1}{K}\sum_{k=1}^{K} s^*_\infty(\mathcal{S}, u, w^k). \tag{14}$$

#### 4.6.2.2 Representation on a Budget using the R2-Indicator

We can now reformulate Problem 1 for the case of the R2-indicator as the performance indicator. We assume throughout this and the following sections that a representation of $\mathcal{Y}_{\text{Par}}$ with $T$ distinct points exists, i.e., that $\mathcal{Y}_{\text{Par}}$ contains at least $T$ distinct points.

▶ Problem 2. [R2-representation on a budget] Suppose we are *given* an instance of problem (8) with unknown efficient set and unknown Pareto front, a reference point $u \in \mathbb{R}^q$ such that $u \geqq f(x)$ for all $x \in \mathcal{X}_{\text{eff}}$, a finite sample of (uniformly distributed) weight vectors $\mathcal{W} \in \mathbb{R}^q_{\geqslant}$, and a fixed budget of representative points $T$ with $T \leq K$ for number of weights $K$. Then, our problem is to *find* a representation $\mathcal{R} \subseteq \mathcal{Y}_{\text{Par}}$ of the Pareto front with cardinality $|\mathcal{R}| = T$, i.e., find a subset $\mathcal{R}$ of the Pareto front containing $T$ distinct nondominated points, that is optimal with respect to the R2-indicator.

Formally, Problem 2 can be written as

$$\min_{\mathcal{R}\subset f(\mathcal{X}),\,|\mathcal{R}|=T} \sum_{k=1}^{K} \min_{f(x^t)\in\mathcal{R}} \left( \max_{j\in\{1,\dots,q\}} w_j^k(u_j - f_j(x^t))\right). \tag{15}$$

Note that we can omit the constant factor $\frac{1}{K}$ from (14) here since it has no impact on the optimal representation $\mathcal{R}$. Note also that the weighted Chebyshev formulation guarantees that all solutions in $\mathcal{R}$ are at least weakly efficient. To guarantee that $\mathcal{R} \subseteq f(\mathcal{X}_{\text{eff}})$ we may add an augmentation term to the weighted Chebyshev formulation

$$\min_{\mathcal{R}\subset f(\mathcal{X}),\,|\mathcal{R}|=T} \sum_{k=1}^{K} \min_{f(x^t)\in\mathcal{R}} \left[ \left( \max_{j\in\{1,\dots,q\}} w_j^k(u_j - f_j(x^t))\right) - \delta \sum_{j=1}^{q} f_j(x^t)\right] \tag{16}$$

with a sufficiently small constant $\delta > 0$. See [6] for an analysis on reasonable choices for this augmentation parameter.

#### 4.6.2.3 Mixed-Integer Reformulation

Problems (15) and (16) both have a min-max-structure which may not be preferable. However, it is possible to model a mixed-integer reformulation.

Indeed, an optimal representation $\mathcal{R}$ for the R2-representation problem is always associated with a (not necessarily unique) assignment between weight vectors $w^k \in \mathcal{W}$ and representative points $f(x^t) \in \mathcal{R}$. For every weight vector $w^k \in \mathcal{W}$ we can identify a point $f(x^t) \in \mathcal{R}$ such that $f(x^t)$ minimizes the $w^k$-weighted Chebyshev distance to the reference point among points in $\mathcal{R}$. We can thus reformulate the R2-representation problem (15) (and similarly its augmented variant (16)) as a mixed integer programming problem by introducing binary variables $z_{tk} \in \{0,1\}$, $t = 1,\dots,T$, $k = 1,\dots,K$ that represent such an optimal assignment. More precisely, $z_{tk} = 1$ if $f(x^t)$ is *assigned* to weight $w^k$ (and hence minimizes the $w^k$-weighted Chebyshev distance to $u$) and $z_{tk} = 0$ otherwise. If $z_{tk} = 1$ we will also say that the weight $w^k$ is *covered* by the solution $x^t$ (or the point $f(x^t)$). To ensure that each weight $w^k \in \mathcal{W}$ is actually covered by a point $f(x^t) \in \mathcal{R}$, we enforce that $\sum_{t=1}^{T} z_{tk} = 1$ for all $k = 1,\dots,K$. Similarly, we generally want to avoid unnecessary points in the set $\mathcal{R}$ and thus ensure that every solution covers at least one weight by requiring $\sum_{k=1}^{K} z_{tk} \geq 1$ for all $t = 1,\dots,T$.

This approach is combined with the standard reformulation of (weighted) Chebyshev distances using auxiliary upper bound variables that are minimized in the objective function. Towards this end, let $k \in \{1,\dots,K\}$ and $t \in \{1,\dots,T\}$ be fixed and consider the subproblem of choosing the solution $x^t \in \mathcal{X}$ such that $f(x^t)$ minimizes the $w^k$-weighted Chebyshev distance from the reference point $u$:

$$\min_{x^t\in\mathcal{X}} \max_{j=1,\dots,q} w_j^k\,(u_j - f_j(x^t)). \tag{17}$$

Let $d_{tk} \geq 0$ be an additional continuous variable that is supposed be equal to the value of the $w^k$-weighted Chebyshev distance between $f(x^t)$ and $u$ at optimality. Then (17) is equivalent to

$$\begin{aligned} \min\quad & d_{tk} \\ \text{s.t.}\quad & w_j^k(u_j - f_j(x^t)) \leq d_{tk}, \quad \forall\, j = 1,\dots,q \\ & x_t \in \mathcal{X},\ d_{tk} \geq 0. \end{aligned}$$

Combining this reformulation with the assignment variables introduced above yields the following integer programming formulation for (15):

$$\min \quad \sum_{t=1}^{T}\sum_{k=1}^{K} d_{tk} \tag{18}$$

$$\text{s.t.} \quad w_j^k(u_j - f_j(x^t)) \leq d_{tk} + (1 - z_{tk}) \underbrace{w_j^k u_j}_{=:M_{jk}} \quad \forall\, j=1,\dots,q,\ t=1,\dots,T,\ k=1,\dots,K, \tag{19}$$

$$\sum_{t=1}^{T} z_{tk} = 1 \qquad \forall\, k = 1,\dots,K, \tag{20}$$

$$\sum_{k=1}^{K} z_{tk} \geq 1 \qquad \forall\, t = 1,\dots,T, \tag{21}$$

$$x_t \in \mathcal{X},\quad z_{tk} \in \{0,1\},\ d_{tk} \geq 0 \qquad \forall t = 1,\dots,T,\ k = 1,\dots,K. \tag{22}$$

Note that constraints (19) are inactive whenever $z_{tk} = 0$ and $f_j(x^t) \geq 0$ for all $j = 1,\dots,q$, which is satisfied at optimality under our assumption that the Pareto front is a subset of the nonnegative orthant in (9). Indeed, if $z_{tk} = 0$ then (19) is equivalent to $-f_j(x^t) \leq d_{tk}$ which is always satisfied with the smallest possible value of $d_{tk} = 0$ when $f_j(x^t) \geq 0$. Note that dominated solutions $x \notin \mathcal{X}_{\text{eff}}$ may get penalized by this reformulation when $z_{tk} = 0$. This is, however, irrelevant for the optimal solution. Thus, $M_{jk} = w_j^k u_j$ is a sufficiently large constant to be used in a "big-M" constraint in this model. Moreover, the choice of $M_{jk} = w_j^k u_j$ allows for a simplification of constraints (19) to

$$w_j^k(z_{tk}u_j - f_j(x^t)) \leq d_{tk}, \quad \forall\, j = 1,\dots,q,\ t = 1,\dots,T,\ k = 1,\dots,K. \tag{19$'$}$$

Note also that an augmentation term can be added to the objective function (18) by replacing it by

$$\sum_{t=1}^{T}\sum_{k=1}^{K}\left(d_{tk} + z_{tk}\cdot\delta\sum_{j=1}^{q} f_j(x^t)\right), \tag{18$'$}$$

where $\delta > 0$ is a sufficiently small constant (c.f. formulation (16) above). This yields, however, nonlinear terms in the objective function even if the original problem (8) is a linear problem.

The complexity of the MIP formulation (18)-(22) depends on the type of considered multiobjective optimization problem (8). We will discuss several interesting special cases in the following subsections. The MIP formulation contains $T \cdot K$ binary $z$-variables, $T \cdot K$ continuous $d$-variables, and $T$ solution vectors $x$ that must be feasible for the original problem (8). Besides these feasibility constraints, the formulation contains $q \cdot T \cdot K$ bound constraints (19′) and $K + T$ assignment constraints (20) and (21).

#### 4.6.2.4 Multiobjective (Mixed-Integer) Linear Problems

Consider the case that problem (8) is a multiobjective linear programming problem

$$\begin{aligned} \min \quad & Cx \\ \text{s.t.} \quad & Ax = b \\ & x \geqq 0 \end{aligned} \tag{23}$$

with a rational objective matrix $C \in \mathbb{Q}^{q\times n}$, a rational constraint matrix $A \in \mathbb{Q}^{m\times n}$, and a rational right-hand-side vector $b \in \mathbb{Q}^m$. We assume that $\text{rank}(A) = m \leq n$ and that $\mathcal{X}_{\text{eff}} \neq \emptyset$ and bounded. Then the MIP formulation (18)-(22) yields a mixed integer linear programming problem

$$\min \quad \sum_{t=1}^{T}\sum_{k=1}^{K} d_{tk} \tag{24}$$

$$\text{s.t.} \quad w_j^k(z_{tk}u_j - C_{j\bullet}x^t) \leq d_{tk}, \qquad \forall\, j=1,\dots,q,\ t=1,\dots,T,\ k=1,\dots,K, \tag{25}$$

$$\sum_{t=1}^{T} z_{tk} = 1 \qquad \forall\, k=1,\dots,K, \tag{26}$$

$$\sum_{k=1}^{K} z_{tk} \geq 1 \qquad \forall\, t=1,\dots,T, \tag{27}$$

$$Ax^t = b,\ x^t \geqq 0,\ z_{tk} \in \{0,1\},\ d_{tk} \geq 0 \qquad \forall t=1,\dots,T,\ k=1,\dots,K. \tag{28}$$

The MILP (24)-(28) can be solved with available solvers as, e.g., CPLEX or Gurobi. This is also possible when the original problem (8) is a multiobjective mixed-integer linear programming problem, i.e., if some of the original variables in $x \in \mathcal{X}$ have to satisfy integrality constraints.

### 4.6.3 Illustrative Example

We have implemented the MILP model (24)-(28) for a biobjective binary knapsack problem in AMPL and solved it with Gurobi 10.0.2. Figure 9 shows R2-optimal representations $\mathcal{R}$ for an instance of a biobjective binary knapsack problem with 30 items together with the complete Pareto front $\mathcal{Y}_{\text{Par}}$. We chose $T \in \{3,4,5\}$ and considered a weight set $\mathcal{W}$ given by

$$\mathcal{W} = \left\{ \left( \frac{i-1}{K-1}, \frac{K-i}{K-1} \right) : i = 1,\dots,K \right\}$$

with $K = 11$ for all shown representations. The reference point $u$ is set to the ideal point, shifted by a multiplicative factor of 1.01.

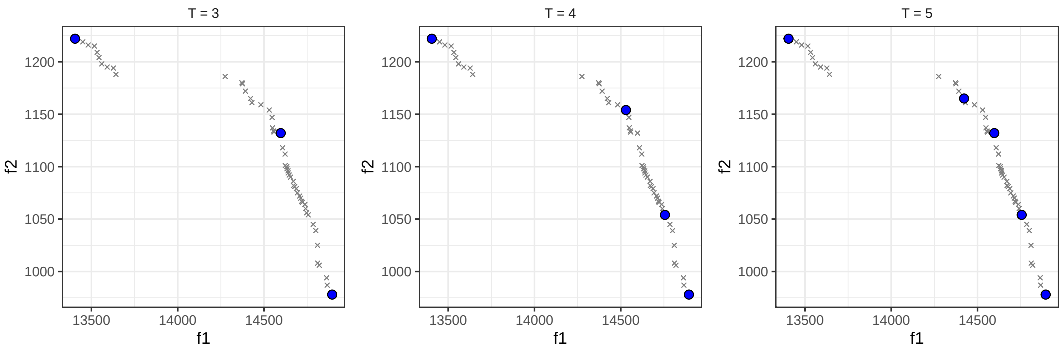



**Figure 9** Optimal R2-representations for an instance of a bi-objective binary knapsack problem with 30 items and with $T = 3$ (left), $T = 4$ (center), and $T = 5$ (right) points. In each case, the returned representation is indicated by blue circles. The set of weight vectors $\mathcal{W}$ was uniformly generated such that $\left(\frac{i-1}{K-1}, \frac{K-i}{K-1}\right)$, $i \in \{1,\dots,K\}$, and $K = 11$.

### 4.6.4 Concluding Remarks and Future Research

In this work, we provide a closed form mixed integer programming formulation for the computation of (globally) optimal R2-representations for general multiobjective optimization problems. The complexity of the formulation depends on the underlying multiobjective optimization problem. For example, if the underlying problem is a multiobjective linear or mixed integer linear programming problem, then our formulation is a mixed integer linear programming problem that can be solved by available solvers.

Future research should discuss multiobjective problems with a more complex structure or with an increasing number of objective functions. In this case, the development of efficient solution heuristics, including iterative and greedy approaches, could be complemented by further improvements of the presented mixed integer programming formulations.

Given that, as it was mentioned above, the R2-indicator becomes equivalent to hypervolume under appropriate settings [7, 22], our formulation could probably also be adapted to finding an approximately optimal hypervolume representation. The representation would be only approximately optimal since we use a finite number of weight vectors. This would require adaptation of our model to a slightly different, inverse version of the Chebyshev function.

In this report, we focus on the all-at-once approach, which may be difficult for available solvers. If the model is too difficult for available solvers, this approach could also be easily adapted to locating the solutions one-by-one in a greedy manner. We would just need to treat values of already selected solutions as fixed parameters and optimize location(s) of just a single (or several) new solution(s). Since R2 most likely shares with the hypervolume the property of being a non-decreasing submodular function, the greedy approach would probably give some approximation guarantee. Another approach that could be used if the model is too difficult for available solvers would be to solve it with some single-objective metaheuristics or a hybrid approach combining metaheuristics with solvers called for some smaller subproblems. Furthermore, even if the proposed model could not be solved for some practical problems in acceptable time, it could still be applied to generate benchmarks for benchmarking heuristic methods aiming at finding a given number of efficient solutions.

### 4.6.5 Acknowledgements

This work has benefited substantially from Dagstuhl Seminar 23361: Multiobjective Optimization on a Budget, and from wider discussions with Matthias Ehrgott within the working group on representations. S. R. Hunter gratefully acknowledges support from the United States Air Force Office of Scientific Research under award number FA9550-23-1-0488.

#### References

1 C. Audet, J. Bigeon, D. Cartier, S. Le Digabel, and L. Salomon. Performance indicators in multiobjective optimization. *European Journal of Operational Research*, 292:397–422, 2021.
2 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.
3 B. Bozkurt, J. W. Fowler, E. S. Gel, B. Kim, M. Koksalan, and J. Wallenius. Quantitative comparison of approximate solution sets for multicriteria optimization problems with weighted tchebycheff preference function. *Operations Research*, 58(3):650–659, 2010.
4 D. Brockhoff, T. Wagner, and H. Trautmann. R2 indicator-based multiobjective search. *Evolutionary Computation*, 23(3):369 – 395, 2015.

5 G. Ceyhan, M. Köksalan, and B. Lokman. Finding a representative nondominated set for multi-objective mixed integer programs. *European Journal of Operational Research*, 272(1):61–77, 2019.

6 K. Dächert, J. Gorski, and K. Klamroth. An augmented weighted Tchebycheff method with adaptively chosen parameters for discrete bicriteria optimization problems. *Computers & Operations Research*, 39:2929–2943, 2012.

7 J. Deng and Q. Zhang. Approximating hypervolume and hypervolume contributions using polar coordinate. *IEEE Transactions on Evolutionary Computation*, 23(5):913–918, 2019.

8 M. Ehrgott. *Multicriteria Optimization*. Springer, 2005. Second edition.

9 M. Ehrgott, L. Shao, and A. Schöbel. An approximation algorithm for convex multi-objective programming problems. *Journal of Global Optimization*, 50:397–416, 2011.

10 A. P. Guerreiro, K. Klamroth, and C. M. Fonseca. Theoretical aspects of subset selection in multi-objective optimisation. In D. Brockhoff, M. Emmerich, B. Naujoks, and R. Purshouse, editors, *Many-Criteria Optimization and Decision Analysis*, Natural Computing Series, pages 213–239. Springer, Cham, 2023.

11 H. W. Hamacher, C. R. Pedersen, and S. Ruzika. Finding representative systems for discrete bicriterion optimization problems. *Operations Research Letters*, 35(3):336–344, 2007.

12 M. P. Hansen and A. Jaszkiewicz. Evaluating the quality of approximations of the nondominated set. *Institute of Mathematical Modeling, Technical University of Denmark*, IMM Technical Report IMM-REP-1998-7, 1998.

13 M. Li and X. Yao. Quality evaluation of solution sets in multiobjective optimisation: A survey. *ACM Computing Surveys*, 52(2), 2019.

14 M. Masin and Y. Bukchin. Diversity maximization approach for multiobjective optimization. *Operations Research*, 56(2):411–424, 2008.

15 K. Miettinen. *Nonlinear Multiobjective Optimization*. Springer, 1998.

16 L. Paquete, B. Schulze, M. Stiglmayr, and A. C. Lourenço. Computing representations using hypervolume scalarizations. *Computers & Operations Research*, 137:105349, 2022.

17 H. Reuter. An approximation method for the efficiency set of multiobjective programming problems. *Optimization*, 21(6):905–911, 1990.

18 S. Sayin. Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming. *Mathematical Programming*, 87:543–560, 2000.

19 S. Sayin. A procedure to find discrete representations of the efficient set with specified coverage errors. *Operations Research*, 51(3):427–436, 2003.

20 S. Sayin and P. Kouvelis. The multiobjective discrete optimization problem: A weighted min-max two-stage optimization approach and a bicriteria algorithm. *Management Science*, 51(10):1572–1581, 2005.

21 B. Schulze. *New Perspectives on Multiobjective Knapsack Problems*. PhD thesis, University of Wuppertal, 2017.

22 K. Shang, H. Ishibuchi, and X. Ni. R2-based hypervolume contribution approximation. *IEEE Transactions on Evolutionary Computation*, 24(1):185–192, 2020.

23 R. E. Steuer. *Multiple Criteria Optimization: Theory, Computation, and Application*. Wiley Series in Probability and Mathematical Statistics. Wiley, 1986.

24 J. Sylva and A. Crema. A method for finding well-dispersed subsets of non-dominated vectors for multiple objective mixed integer linear programs. *European Journal of Operational Research*, 180(3):1011–1027, 2007.

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

## 4.7 Modeling and Decomposition – Biobjective Block-Coordinate Descent

*Andrea Raith (University of Auckland, NZ), Kerstin Dächert (HTW Dresden, DE), Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR), Joshua D. Knowles (Schlumberger Cambridge Research, GB), Aneta Neumann (University of Adelaide, AU), Frank Neumann (University of Adelaide, AU), Anita Schöbel (Fraunhofer ITWM – Kaiserslautern, DE), and Margaret M. Wiecek (Clemson University, US)*



### 4.7.1 Introduction

Many real-world decision problems are too large or too complex to be modeled and solved as one large optimization problem. A well established idea to solve such problems is to decompose them into smaller subproblems that can be solved so that their optimal solutions recover an optimal solution to the original problem. In the literature, the approaches to decomposing the original problem and to solving the subproblems are designed in many different ways, which we now briefly review. Decomposition of a complex problem into subproblems and coordination of their solutions are meant to bring savings and are therefore often used in optimization-based decision processes with scarce resources or limited budgets.

The original or overall optimization problem models a complex decision-making process on a man-made system whose performance is determined by the values of the decision variables, that is subject to constraint functions, and is evaluated by one or more objective functions. We refer to the original or overall optimization problem as an All-in-One (AiO) problem. Decomposition of the AiO problem into subproblems may be conducted with respect to the following concepts:

- The disciplines of science or engineering that are used to develop the mathematical model of the AiO system (e.g., control theory, mechanics, mixed-integer programs, PDEs);
- The physical parts the AiO system consists of – the subsystems, components, and subcomponents;
- The structure of the AiO mathematical model that is reflected in the placement of the decision variables in the objective and constraint functions;
- The tradeoffs between specific objectives the decision maker (DM) is able to assess;
- The scenarios in which the AiO system is expected to perform;
- The scalarization type that is applied to multiple objective functions to transform them into a single objective function.

Given the subproblems, various methods for solving them and coordinating the solutions obtained have been proposed with the ultimate goal to retrieve an optimal AiO solution without actually solving the AiO problem. The solution methods are based on the following strategies:

- Simultaneous coordination: the subproblems are solved concurrently and independently with minimum or no information exchanged among them;
- Sequential coordination: the subproblems are solved consecutively or in subsequent stages, and an optimal solution of a predecessor subproblem is carried over to solve the successor subproblem;

- Hierarchical coordination: the subproblems are solved consecutively and an optimal solution of a successor subproblem guarantees the optimality of the predecessor subproblem that has already been achieved.

For more details on the decomposition and coordination the reader is referred to [9, 4, 15] and the references therein.

In this paper we are interested in a decomposition of a decision problem with respect to the mathematical model and in a sequential coordination that allows the construction of an AiO optimal solution. In particular, the decomposition is conducted with respect to the decision variables that make up the model, while the coordination proceeds in a multistage fashion, meaning that subsequent optimizations at the level of subproblems are performed in lower-dimensional spaces and therefore offer computational savings. Many planning problems consist of subproblems that are solved in stages which often (but not always) proceed in a given order. To motivate our focus on this type of complex decision making problems, we provide examples from location science and from public transportation to illustrate that multistage decision problems appear naturally.

In location science, the p-median problem consists of two stages: In the first stage, the $p$ new facilities are determined. In the second stage, the assignment from the demand points to the new facilities is computed. If the facilities are fixed, finding the best assignment is easy. On the other hand, when the assignment is fixed, the facilities follow directly. While each stage is polynomially solvable, finding the facilities for the AiO problem with the assignment is NP-hard.

Public transportation is a complex man-made system usually designed via subproblems that are associated with several stages. First, the location of the stations is determined. Then the lines that connect the stations are constructed. After the lines have been planned, one determines a timetable, and later on, a vehicle- and finally the crew schedule. Consider the three consecutive subproblems: line planning, timetabling, and vehicle scheduling. Although analyzed and solved consecutively in most theoretical and practical approaches, these three subproblems are interconnected. Sometimes one could save a complete vehicle by only one small modification of the lines. However, each of the three subproblems is already NP-hard by itself, so there is no chance of determining an exact optimal solution simultaneously by solving the AiO problem. Hence, these subproblems are solved step by step according to the following strategy:

1. Finding a starting solution: An easy (greedy-type) heuristic to find an initial solution is to proceed sequentially. One first determines a line plan, then adds a timetable, and finally a vehicle schedule. It is well known that this approach is only a heuristic.
2. Improving the starting solution: Given the initial solution, one chooses one of the subproblems and optimizes with respect to its variables while keeping all the other variables fixed. For example, one keeps the line plan and the vehicle schedule fixed, but re-optimizes the timetable. Then another subproblem, e.g., line planning, is selected and re-optimized with respect to its variables. This process is performed until no further improvement is possible. The resulting coordination method is called the *blockwise coordinate descent* (BCD).

Figure 10 depicts a graph (called *Eigenmodel*) which illustrates finding a starting solution as well as the iteration steps (red). Notice that it contains more nodes than the ones referring to the algorithms mentioned before. In particular, each path through the Eigenmodel corresponds to solving a sequence of suproblems and hence to a heuristic. The name *Eigenmodel* refers to the own (=eigen) subproblems which form the nodes of the depicted graph. Eigenmodels have been introduced in [12], and analyzed, e.g., in [8, 14].

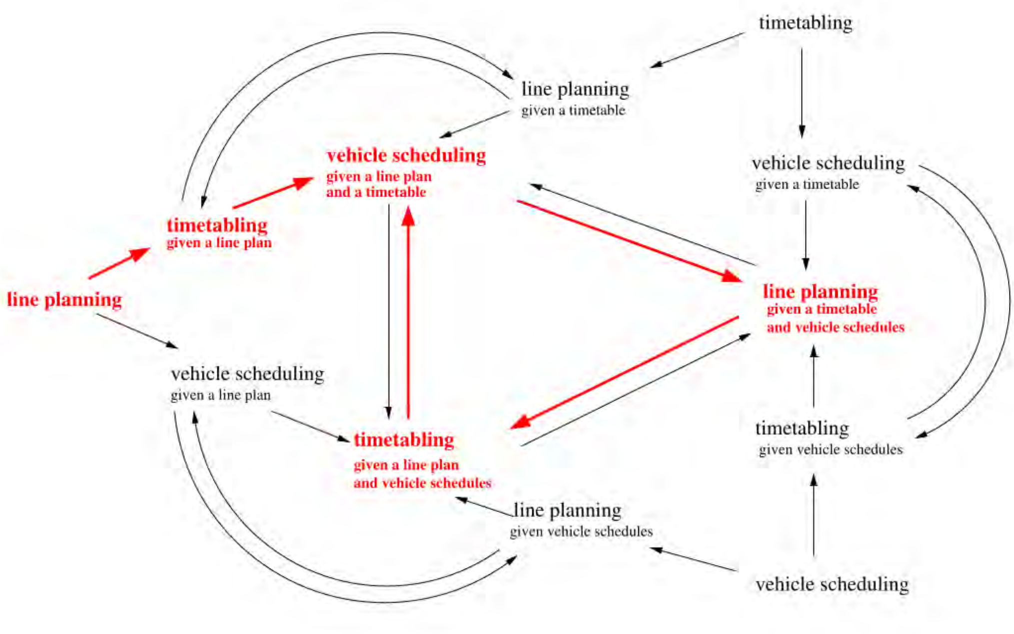



**Figure 10** A part of the Eigenmodel for public transportation planning

In the examples presented, a scalar-valued objective function is associated with each subproblem resulting in multiple and conflicting objectives. As a result, a vector-valued objective function is related to the AiO problem but is ignored by the BCD heuristic that solves the single objective optimization problem at every stage. In this study we recognize the existence of an AiO vector-valued function that is carried by every subproblem at every stage. In other words, we deal with AiO multiobjective optimization problems (MOPs) that remain multiobjective at every stage. MOPs lend themselves to decomposition into multiobjective (or single objective) subproblems whose solutions shall provide the efficient solutions to the original MOPs. In the public transportation example one might follow several objectives: the costs of the public transport system should be minimized. At the same time, the traveling time for passengers as well as the carbon emission should be minimized.

The standard single objective BCD relies on a well-established algorithm with a proof of convergence [1] for continuous optimization problems. The contribution of this preliminary study is the formulation of the BCD for the biobjective two-stage case with some supporting theory and implementation variants. The latter include approximation algorithms and evolutionary heuristics.

### 4.7.2 Block coordinate descent algorithm for two criteria and two blocks

Consider the following biobjective unconstrained optimization problem with two blocks:

$$\begin{aligned} \min\,&[f_1(x_1,x_2), f_2(x_1,x_2)] \\ s.t.\quad &(x_1,x_2)\in\mathbb{R}^2 \end{aligned} \tag{29}$$

where $f_i : \mathbb{R}^2 \to \mathbb{R}$, $i \in \{1,2\}$, and $x_i \in \mathbb{R}$ for $i \in \{1,2\}$ are the two scalar coordinate directions, also referred to as blocks of one or more variables in the general block coordinate descent method. Let

- $E^*$ denote the efficient set of (29)
- $E_i^k$ denote the efficient set in iteration $k$ for block $i$
- $E^k$ denote the efficient set at the completion of iteration $k$
- $E^N$ denote the efficient set at the completion of N iterations
- $\mathit{Eff}(S)$ denote the operator of extracting the efficient points from an arbitrary set S

#### 4.7.2.1 Algorithm: Biobjective Block Coordinate Descent (BBCD)

1. Initialization: starting point $x^0 = (x_1^0, x_2^0) \in \mathbb{R}^2$ (not necessarily an efficient starting solution), $E^0 = \{(x_1^0, x_2^0)\}$, $k := 1$
2. Iteration $k$:

   **(B1)** (fix variable $x_2$)
   For all $\mathbf{y} \in E^{k-1}$: Solve problem (29) with $x_2 = y_2$, i.e.,

   $$E_1^k(\mathbf{y}) = \arg\min_{x_1} [f_1(x_1, y_2), f_2(x_1, y_2)]$$

   Compute $E_1^k = \mathit{Eff}\left(\bigcup_{y \in E^{k-1}} E_1^k(\mathbf{y})\right)$

   **(B2)** (fix variable $x_1$)
   For all $\mathbf{y} \in E_1^k$: Solve problem (29) with $x_1 = y_1$, i.e.,

   $$E_2^k(\mathbf{y}) = \arg\min_{x_2} [f_1(y_1, x_2), f_2(y_1, x_2)]$$

   Compute $E_2^k = \mathit{Eff}\left(\bigcup_{\mathbf{y} \in E_1^k} E_2^k(\mathbf{y})\right)$

   Set $E^k = E_2^k$
3. If $E^k = E^{k-1}$ stop, set $E^N = E^k$, otherwise set $k = k + 1$ and return to step 2.

To understand the properties of the proposed BBCD algorithm, the extension of block coordinate descent to the biobjective case, we seek to answer the following research questions. Let $N$ be the number of iterations of the BBCD algorithm ($N$ can be assumed finite or infinite).

1. Under which conditions does the set $E^N$ contain only efficient solutions, i.e., $E^N \subseteq E^*$?
2. Under which conditions are all efficient solutions found, i.e., $E^* \subseteq E^N$?
3. How big is $N$, or in other words, what is the number of iterations needed to achieve the two goals above?

#### 4.7.2.2 Auxiliary background

The following notation is used: a vector-valued function $\mathbf{f} : \mathbb{R}^2 \mapsto \mathbb{R}^2$, scalar-valued function $f : \mathbb{R}^2 \mapsto \mathbb{R}$, level set $\mathcal{L}_{\leq}(f(\bar{\mathbf{x}})) = \{\mathbf{x} \in \mathbb{R}^2 : f(\mathbf{x}) \leq f(\bar{\mathbf{x}})\}$, and level curve $\mathcal{L}_{=}(f(\bar{\mathbf{x}})) = \{\mathbf{x} \in \mathbb{R}^2 : f(\mathbf{x}) = f(\bar{\mathbf{x}})\}$.

We also define the coordinate lines (hyperplanes) passing through a point $\bar{\mathbf{x}} \in \mathbb{R}^2$ for $i \in \{1, 2\}$: $H_i(\bar{\mathbf{x}}) = \{(x_1, x_2) \in \mathbb{R}^2 : x_j = \bar{x}_j \text{ for } j \in \{1, 2\} \text{ and } j \neq i\}$, i.e. $H_i(\bar{\mathbf{x}})$ represents the coordinate line in the $x_i$-direction.

Theorem 1 shows how level sets and curves can be used to characterize the efficient solutions of an unconstrained $p$-objective optimization problem. Figure 11 depicts two level curves of two hypothetical objective functions indicating that a point $x \in \mathbb{R}^2$ is not efficient.

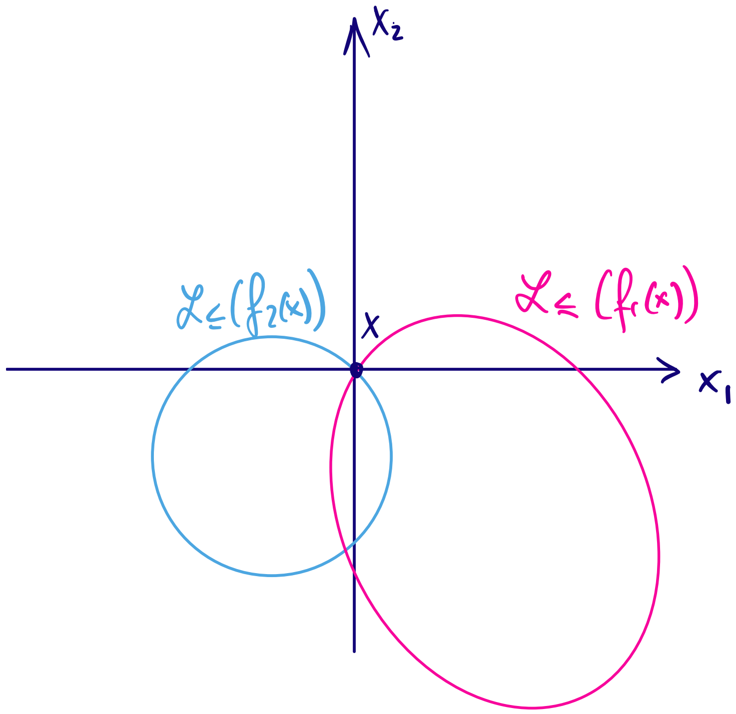



**Figure 11** Case 1: the intersection of $\mathcal{L}_{\leq}(f_1(\mathbf{x}))$ and $\mathcal{L}_{\leq}(f_2(\mathbf{x}))$ is aligned with a search direction.

▶ **Theorem 1.** *[6] Let $\hat{\mathbf{x}} \in \mathbb{R}^n$ and define $\hat{y}_k = f_k(\hat{\mathbf{x}}), k = 1, 2, \ldots, p$. Then $\hat{\mathbf{x}}$ is efficient if and only if*

$$\bigcap_{k=1}^{p} \mathcal{L}_{\leq}(\hat{y}_k) = \bigcap_{k=1}^{p} \mathcal{L}_{=}(\hat{y}_k) \quad . \tag{30}$$

A curve in $\mathbb{R}^2$ is said to be smooth provided for every point on this curve there exists a neighborhood on which this curve is a graph of a differentiable function. In the following we assume $f : \mathbb{R}^2 \mapsto \mathbb{R}$ differentiable and that $\mathcal{L}_{=}(f(\bar{\mathbf{x}}))$ is a smooth curve in $\mathbb{R}^2$, and therefore $\mathcal{L}_{\leq}(f(\hat{\mathbf{x}})) \cap H_i(\bar{\mathbf{x}}) \neq \emptyset$.

#### 4.7.2.3 Conjectures with proof outlines

The first research question asks whether all solutions found by the BBCD are efficient, i.e., $E^N \subseteq E^*$. In the following it is assumed that the functions $f_1, f_2$ are convex and differentiable.

*Proof sketch for $E^N \subseteq E^*$:* Assume $\mathbf{x} \in E^N$ and $\mathbf{x} \notin E^*$. In particular, $E^N = E^k = E^{k-1}$ where $k$ is the number of iterations of the BBCD. According to Theorem 1, an efficient point satisfies (30). Since $\mathbf{x}$ is not efficient, the intersection of the level sets $\mathcal{L}_{\leq}(f_1(\mathbf{x}))$ and $\mathcal{L}_{\leq}(f_2(\mathbf{x}))$ has a nonempty interior. Two cases of this intersection are shown in Figures 11 and 12.

In Case 1 depicted in Figure 11, we observe that the search into one of the search directions leads into the area containing solutions $\tilde{\mathbf{x}}$ that improve both $f_1$ and $f_2$, hence $\mathbf{f}(\tilde{\mathbf{x}}) \leq \mathbf{f}(\mathbf{x})$. In the scenario shown in Figure 11 the BBCD searches into the direction $x_1$ first. We note that in this case, $\mathbf{x}$ remains an efficient solution in the search along $H_1(\mathbf{x})$ as

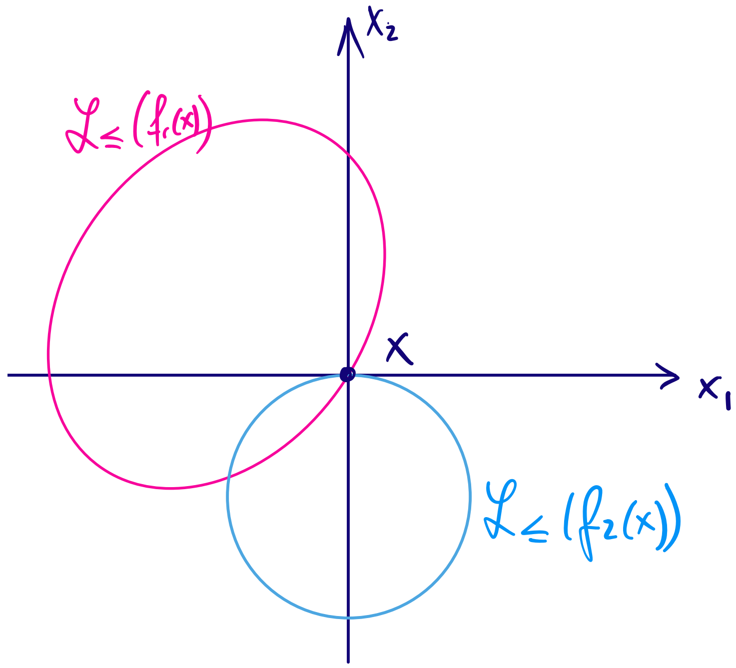



**Figure 12** Case 2: the intersection of $\mathcal{L}_\leq(f_1(\mathbf{x}))$ and $\mathcal{L}_\leq(f_2(\mathbf{x}))$ is *not* aligned with a search direction.

it cannot be dominated by other solutions encountered along $H_1(\mathbf{x})$ (if there are any). (Also the search along $H_1(\mathbf{x})$ yielded at least $\mathbf{x}$ in the last iteration as $E^N = E^k = E^{k-1}$). Once the BBCD searches in direction $x_2$, solutions in the interior of $\mathcal{L}_\leq(f_1(\mathbf{x})) \cap \mathcal{L}_\leq(f_2(\mathbf{x}))$ are encountered. Since $\mathbf{f}(\mathbf{y}) \leq \mathbf{f}(\mathbf{x})$ for any such solution $\mathbf{y} \in \mathcal{L}_\leq(f_1(\mathbf{x})) \cap \mathcal{L}_\leq(f_2(\mathbf{x}))$ we have a contradiction to $\mathbf{x} \in E^N$. Similarly, such a solution $\mathbf{y}$ would be encountered immediately if $H_1(\mathbf{x})$ intersects the interior of $\mathcal{L}_\leq(f_1(\mathbf{x})) \cap \mathcal{L}_\leq(f_2(\mathbf{x}))$.

In Case 2 depicted in Figure 12, the interior of $\mathcal{L}_\leq(f_1(\mathbf{x})) \cap \mathcal{L}_\leq(f_2(\mathbf{x}))$ and $H_i(\mathbf{x})$ do not intersect for $i = 1, 2$. Here, we know that, starting from $\mathbf{x}$, we can find efficient solutions along either $H_1(\mathbf{x})$ or $H_2(\mathbf{x})$ ($H_1(\mathbf{x})$ in the scenario shown in Figure 12). (Otherwise $\mathbf{x}$ would be efficient as the level sets do not intersect). Moving at least some way in a direction that improves an objective ($f_1$ in Figure 12) would yield other efficient solutions $\mathbf{y}$ such that $\mathbf{f}(\mathbf{y})$ and $\mathbf{f}(\mathbf{x})$ do not dominate each other.
The point $\mathbf{y}$ is shown in Figure 13 together with the level sets of $\mathbf{y}$ where we have $\mathcal{L}_\leq(f_1(\mathbf{y})) \subseteq \mathcal{L}_\leq(f_1(\mathbf{x}))$ and $\mathcal{L}_\leq(f_2(\mathbf{x})) \subseteq \mathcal{L}_\leq(f_2(\mathbf{y}))$. The search along the other block $H_2(\mathbf{y})$ from point $\mathbf{y}$ will move towards and into the intersection $\mathcal{L}_\leq(f_1(\mathbf{x})) \cap \mathcal{L}_\leq(f_2(\mathbf{x}))$ as $f_2$ decreases and $f_1$ increases. We have $\mathbf{f}(\tilde{x}) \leq \mathbf{f}(\mathbf{x})$ for any solution $\tilde{\mathbf{x}} \in H_2(\mathbf{y}) \cap \mathcal{L}_\leq(f_1(\mathbf{x})) \cap \mathcal{L}_\leq(f_2(\mathbf{x}))$. If $\tilde{x}$ itself is not efficient as there exists another $\hat{x}$ constructed by the BBCD with $\mathbf{f}(\hat{x}) \leq \mathbf{f}(\tilde{x})$, then we also have $\mathbf{f}(\hat{x}) \leq \mathbf{f}(\mathbf{x})$. This is a contradiction to $x \in E^N$.

The second research questions investigates whether all efficient solutions can be found by the BBCD. To address the question whether $E^* \subseteq E^N$, the objective functions are assumed to be strictly convex (to avoid weakly efficient solutions). For the inclusion of interest, the differentiability is required to avoid stalling as happens in the following example (31) where

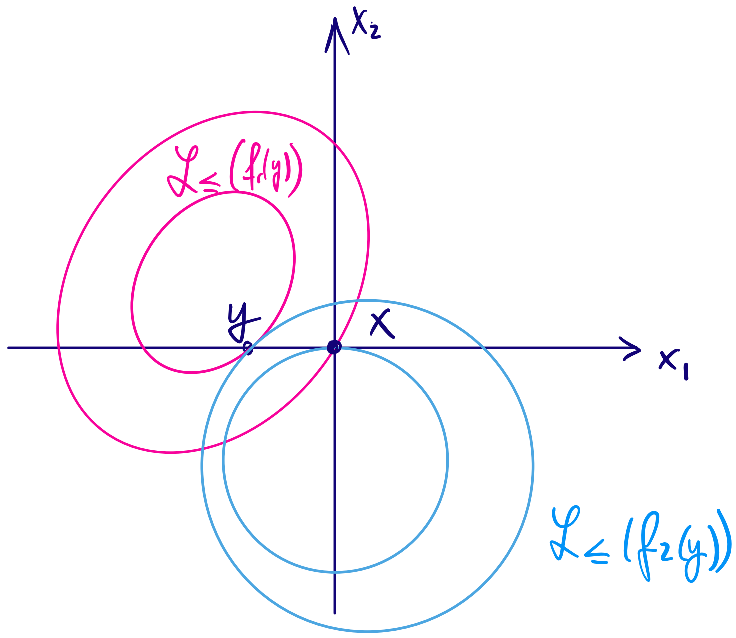



**Figure 13** Case 2 point $\mathbf{y}$ and associated level sets.

the BBCD is unable to identify the complete $E^*$.

$$\begin{aligned} &\min\,[x_1 + 2|x_2|, -x_1 + 2|x_2|] \\ &\qquad s.t. \quad (x_1, x_2) \in \mathbb{R}^2 \end{aligned} \tag{31}$$

In this example, the efficient set is $E^* = \{(x_1, x_2) \in \mathbb{R}^2 | x_1 \in \mathbb{R}, x_2 = 0\}$.

*Proof idea for $E^* \subseteq E^N$:* Start with an efficient solution in $E^N$, and then demonstrate that the search will always expand outwards along connected efficient solutions until the lexicographic solutions are reached (and hence everything in between). This argument should work as the efficient set is connected for convex differentiable multiobjective optimization problems [11].

The third research question addresses the number of required iterations, i.e,. what is $N$ equal to? A first exploration of this question suggests the following observations. Firstly, we observe that $N = 1$ in the case of a simple quadratic problem. Secondly, based on a nonlinear and convex example problem where $E^*$ is unbounded, an infinite number of iterations is required.

#### 4.7.2.4 Ongoing and future work

The initial investigation presented in this section proposed the BBCD for a problem with two variables, where two of the research questions (the inclusion $E^* \subseteq E^N$ and the number of iterations) remain open. This theory needs to be further extended to the case of $n$ coordinate directions $x_1, x_2, \ldots, x_n$, or blocks of variables $\mathbf{x}_i$, as well as $p > 2$ objectives.

We recognize that the block coordinate descent is a powerful heuristic tool to solve large-scale practical optimization problems. Further into the future, the proposed BBCD could be similarly applied as a heuristic for solving biobjective (mixed) integer linear programs.

In order to find the most effective implementation in terms of a budget on computational resources or time, the analysis could be performed in two directions. First, the time to solve AiO problems shall be compared to the time to solve them with the BBCD. Second, working only with the BBCD, the time spent on balancing between solving individual subproblems and exchanging information between them needs to be analyzed.

#### 4.7.3 Multiobjective Evolutionary Algorithms

As alluded to above, the block coordinate descent approach is employed in various applications of optimization such as transportation systems [17], wireless communication networks [18], signal processing [13], multiclass classification [2], and others. We aim to highlight through runtime analysis the benefit of incorporating the block coordinate approach into evolutionary multiobjective algorithms. We employ state-of-the-art algorithms such as the Global Simple Evolutionary Multiobjective Optimizer (GSEMO) [7] and the Nondominated Sorting Genetic Algorithm II (NSGA-II) [3].

We plan to use problems that we understand and analyze the runtime for GSEMO/NSGA-II when incorporating the block coordinate approach. Our aim is to provide an example where we can prove that the block coordinate descent incorporated into these evolutionary algorithms leads to an asymptotic speed up of the optimization process.

##### 4.7.3.1 Algorithms with different levels of complexity

We consider how to implement the algorithm using three different “complexity" algorithms:

- At each iteration, choose only *one* solution to continue
- At each iteration, find *all* the locally Pareto optimal solutions and continue expanding this set (similar to the Pareto Local Search)
- At each iteration, find only $p > 1$ solutions to continue with, using something analogous to the ‘uncrowded hypervolume’ method.

We will compare with Pareto Local Search and multiobjective Branch-and-Bound algorithms to obtain or reuse results on how to continually expand a set to approximate a Pareto set. It might be possible to interpret the three different choices of implementation as budgets or relatable to budgets.

##### 4.7.3.2 Candidate Problems

We consider the following problem candidates:

- MO-version HIFF function [16]
- MO-version of bilinear functions [10]
- MO version of $\text{wCLOB}_{l,k}$ (concatenated LeadingOnes with blocks and weights) [5].

Possibilities for allocating budgets may be:

- Computing the whole set of Pareto optimal solutions for a fixed set of search points
- Allocating budgets for solving the different subproblems.

**Algorithm 3:** Global simple evolutionary multiobjective optimizer (GSEMO).

Initialize $x \in \{0,1\}^n$ uniformly at random;
$P \leftarrow \{x\}$;
**repeat**
  Choose $x \in P$ uniformly at random;
  Create $y$ from $x$ by mutation;
  **if** $\nexists w \in P : w \succ y$ **then**
    $P \leftarrow (P \setminus \{z \in P \mid y \succeq z\}) \cup \{y\}$;
**until** *stop*;

#### 4.7.3.3 Runtime Analysis for Block Coordinate MOEAs

We now define a problem where we hope to show the benefit of incorporating the block coordinate approach. We consider a bi-objective version of the $\text{wCLOB}_{l,k}$ (concatenated LeadingOnes with blocks and weights) problem [5].

Our aim is to show a lower bound for the standard GSEMO (see Algorithm 3) and an upper bound for BC-GSEMO (see Algorithm 4), where the lower bound is lower than the upper bound. This would prove a clear advantage of BC-GSEMO over GSEMO for the problems considered. Let $x = (x_{B_1}, \ldots, x_{B_k})$ where $x_{Bi}$ is the ith block of the bitstring $x$ of length $\ell = n/k$.

We define

$$LO_z(x_{B_i}) = \sum_{i=1}^{\ell} \prod_{j=1}^{i} (z_i = x_i)$$

as the number of leading positions where $x_{B_i}$ agrees with a given string $z$.

Let $z^1 = 1^\ell$ and $z^2 = 1^{\ell-d}0^d$, where $d$ is an appropriate constant.

We consider the biobjective problem $f = (f_1, f_2)\colon \{0,1\}^n \to \mathbb{R}^2$ given as

$$f_1(x) = \sum_{i=1}^{k} (\ell+1)^{k-i} \cdot LO_{z^1}(x_{B_i})$$

$$f_2(x) = \sum_{i=1}^{k} (\ell+1)^{k-i} \cdot LO_{z^2}(x_{B_i}) \quad .$$

We consider optimization using the classical GSEMO algorithm (see Algorithm 3) as well as a block coordinate variant called BC-GSEMO (see Algorithm 4). The algorithm optimizes the blocks sequentially and terminates if the optimization part for the current block does not change the population, i.e., there is no change to the population when executing the repeat loop.

To apply the block coordinate mutation to block $i$, we flip each bit of $x_{B_i}$ independently of the others with probability $1/\ell$. For the local variant called the Block coordinate SEMO (BC-SEMO), one randomly chosen bit is flipped in the chosen block $x_{B_i}$.

We consider different budgets of $t_{max}$, e.g.,

- $t_{max} \leq \ell$ (slow parallel progress, good runtime bounds s)
- $t_{max} = c\ell^2$, $c$ appropriate large constant (optimizes block $(i+1)$ for each individual in the population if population size is constant)
- $t_{max} = c'|P|\ell^2$, $c$ appropriate large constant (optimizes block $(i+1)$ for each individual in the population).

**Algorithm 4:** Block coordinate GSEMO (BC-GSEMO).

Initialize $x \in \{0,1\}^n$ uniformly at random;
$P \leftarrow \{x\}$;
$i \leftarrow 0$;
**repeat**
    $t \leftarrow 0$;
    **repeat**
        $t \leftarrow t + 1$;
        Choose $x \in P$ uniformly at random;
        Create $y$ from $x$ by block-coordinate mutation on block $x_{i+1}$;
        **if** $\nexists w \in P : w \succ y$ **then**
            $P \leftarrow (P \setminus \{z \in P \mid y \succeq z\}) \cup \{y\}$;
    **until** $t \geq t_{max}$;
    $i \leftarrow (i+1) \mod k$;
**until** *stop*;

Note that a careful consideration is required for different choices of $t_{max}$, which is potentially dependent on the current population size.

Let $X^*$ be the set of all search points $x$ for which $x_{B_i} = 1^{\ell-d}B^d$ where $B \in \{0,1\}$ for all $i \in \{1,\ldots,k\}$. Note that for each search point $z \in \{0,1\}^n \setminus X^*$, there is at least one search point in $X^*$ that strongly dominates $z$. Hence, those search points are not Pareto optimal.

For each $x \in X^*$, we have

$$f_1(x) + f_2(x) = \sum_{i=1}^{k} (\ell+1)^{k-i}(2\ell-1) \quad .$$

Furthermore, let $x, y \in X^*$ be two search points with $x \neq y$. Then we have $f_1(x) \neq f_1(y)$ and $f_2(x) \neq f_2(y)$. This implies that each search point $x \in X^*$ is Pareto optimal and the Pareto front is given by $f(X^*) = \bigcup_{x \in X^*} f(x)$.

We provided this problem and the algorithmic setting. It remains to show that BC-GSEMO outperforms GSEMO on the problem by providing upper and lower bounds for the algorithms that show the conjectured difference in performance. During the seminar, we worked on proving a conjecture, and believe we may have a result showing an asymptotic advantage to BC-GSEMO based on the application of modern drift theory. The result follows from differences in the evolving population size in the optimization phases of each algorithm, and different interactions between solutions, with BC-GSEMO suffering from less negative drift.

Our goal is also to complement the theoretical analysis with an experimental study on problem instances of realistic size and examine the performance of the two algorithms in terms of the dependence of the input size, and when different types of budget are imposed.

### References


1 D.P. Bertsekas. *Nonlinear Programming.* Athena Scientfic, Belmont, Massachusetts, 2016.

2 Mathieu Blondel, Kazuhiro Seki, and Kuniaki Uehara. Block coordinate descent algorithms for large-scale sparse multiclass classification. *Machine Learning*, 93(1):31–52, 2013.

3 Kalyanmoy Deb, Samir Agrawal, Amrit Pratap, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

**4** T. Dietz, K. Klamroth, K. Kraus, S. Ruzika, L. Schäfer, B. Schulze, M. Stiglmayr, and M. M. Wiecek. Introducing multiobjective complex systems. *European Journal of Operational Research*, 280:581–596, 2020.

**5** Benjamin Doerr, Dirk Sudholt, and Carsten Witt. When do evolutionary algorithms optimize separable functions in parallel? In *FOGA*, pages 51–64. ACM, 2013.

**6** M. Ehrgott. *Multicriteria Optimization*. Springer, second edition, 2005.

**7** Oliver Giel. Expected runtimes of a simple multi-objective evolutionary algorithm. In *IEEE Congress on Evolutionary Computation*, pages 1918–1925. IEEE, 2003.

**8** S. Jäger and A. Schöbel. The blockwise coordinate descent method for integer programs. *Mathematical Methods of Operations Research*, 91:357–381, 2020.

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

**10** Per Kristian Lehre. Runtime analysis of competitive co-evolutionary algorithms for maximin optimisation of a bilinear function. In *GECCO*, pages 1408–1416. ACM, 2022.

**11** K. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, Boston, 1999.

**12** A. Schöbel. An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. *Transportation Research C*, 74:348–365, 2017.

**13** Dongyuan Shi, Bhan Lam, Woon-Seng Gan, and Shulin Wen. Block coordinate descent based algorithm for computational complexity reduction in multichannel active noise control system. *Mechanical Systems and Signal Processing*, 151:107346, 2021.

**14** P. Schiewe and A. Schöbel. Integrated optimization of sequential processes: General analysis and application to public transport. *EURO Journal on Transportation and Logistics*, 11:100073, 2022.

**15** M.M. Wiecek and P. de Castro. Developing decomposition-coordination techniques for many-objective tradeoff analysis. In S. Greco, V. Mousseau, J. Stefanowski, and C. Zopounidis, editors, *Intelligent Decision Support Systems – Combining Operations Research and Artificial Intelligence*, Springer Series on MCDM, pages 307–329. Springer, 2022.

**16** Richard A. Watson, Gregory Hornby, and Jordan B. Pollack. Modeling building-block interdependency. In *PPSN*, volume 1498 of *Lecture Notes in Computer Science*, pages 97–108. Springer, 1998.

**17** Zewen Wang, Kai Zhang, Xinyuan Chen, Meng Wang, Renwei Liu, and Zhiyuan Liu. An improved parallel block coordinate descent method for the distributed computing of traffic assignment problem. *Transportmetrica A: Transport Science*, 18(3):1376–1400, 2022.

**18** Qingqing Wu, Yong Zeng, and Rui Zhang. Joint trajectory and communication design for multi-uav enabled wireless networks. *IEEE Trans. Wirel. Commun.*, 17(3):2109–2121, 2018.

## 4.8 Exploring correlations in multi-objective optimization

*Tea Tusar (Jozef Stefan Institute – Ljubljana, SI), Dimo Brockhoff (INRIA Saclay – Palaiseau, FR), Jonathan Fieldsend (University of Exeter, GB), Boris Naujoks (TH Köln, DE), Alma Rahat (Swansea University, GB), Vanessa Volz (modl.ai – Copenhagen, DK), Hao Wang (Leiden University, NL), Kaifeng Yang (Univ. of Applied Sciences – Hagenberg, AT)*



### 4.8.1 Introduction

A frequent assumption in evolutionary computation is that all function evaluations take the same amount of time. However, this rarely holds for real-world optimization problems, especially those that rely on simulations for evaluating solutions. There, the evaluation time can differ for different objectives as well as for different solutions.

The case where evaluation time depends on objectives has already been explored in a previous Dagstuhl Seminar [7]. This typically occurs in problems where some objectives can be computed with a closed-form expression while others require lengthy simulations. Various strategies for handling objectives with heterogeneous evaluation times are reviewed in [1].

During this seminar, we focused on the second case, in which the evaluation time depends on solutions. Specifically, we wanted to explore whether the correlation between objectives and their evaluation times can be modeled and exploited to save expensive function evaluations.

### 4.8.2 Motivation from real-world applications

In some real-world problems, the relation between solution properties and evaluation times is rather straightforward. For example, in the tunnel alignment problem [9], where a solution represents a tunnel trajectory, the computational expense of assessing tunnel objectives and constraints is proportional to the length of the tunnel – a longer tunnel will generally take longer to evaluate. Similarly holds for neural architecture search [3], where a solution defines the architecture of a neural network whose training time is strongly positively correlated with its size.

However, there are also other kinds of real-world problems where such a relation is hard to find. Consider the airfoil optimization problem [11], where computational fluid dynamics is used in solution evaluation, and the electrical motor design problem [13], which relies on electromagnetic field simulations. In both cases, evaluation times vary among solutions, but a clear correlation between solution characteristics and evaluation duration has not been discovered.

Another source of solution-dependent evaluation times is the presence of hidden constraints. For instance, the MarioGAN optimization problem [14] involves generating Mario game levels, which are assessed through playthrough simulations with artificial intelligence players. If a generated level cannot be solved (that is, Mario cannot reach the level end), the simulation would continue endlessly unless terminated. The distance in the search space between feasible solutions that are relatively quick to evaluate and infeasible solutions whose evaluation takes a long time can be very small in such cases.

These examples show that the correlation between objective quality and its evaluation time depends on the problem and the solutions. We can model it by considering the evaluation time as an additional independent objective to be minimized.

### 4.8.3 Visualization of correlations

We use search space visualizations to gain a better understanding of the correlations between objectives. The correlation for each pair of objectives is estimated in different regions of the search space using the Pearson correlation coefficient for a small (local) sample of the search space. The Pearson correlation coefficient measures the linear correlation between two samples' objectives and takes a value between $-1$ (perfect linear anti-correlation) and 1 (perfect linear correlation). A 0 value implies that there is no linear dependency between the objectives. The Pearson correlation coefficient is invariant when the two objectives are shifted and/or scaled.

#### 4.8.3.1 Experimental setup

For demonstration purposes, we choose some continuous test problems with 2-D search spaces that are straightforward to visualize. They have either two, three or five objectives and various characteristics (more details below). We assume minimization of their $m$ objectives.

The 2-D problem search space is discretized into a grid of $501 \times 501$ points. For each grid point $x = (x_1, x_2)$, the correlation between two objectives is computed with the Pearson correlation coefficient as follows. First, $p$ equidistant points are created on the circle with radius $10^{-6}$ centered at $(x_1, x_2)$ with one point placed at position $(x_1 + 10^{-6}, x_2)$, see Figure 14. Next, the $p$ points are evaluated, i.e. $m$ objective values are computed for each of them. Finally, the correlation between each pair of objectives at $x$ is estimated with the corresponding Pearson correlation coefficient for the set of $p$ points. Note that the $p$ points could have been constructed also in some other way. We opted for this deterministic approach to minimize the disturbances caused by a stochastic choice of point placement. In all experiments, the number of points $p$ was set to 100.

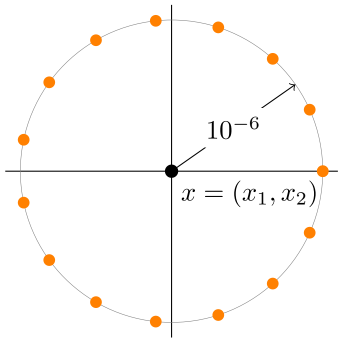
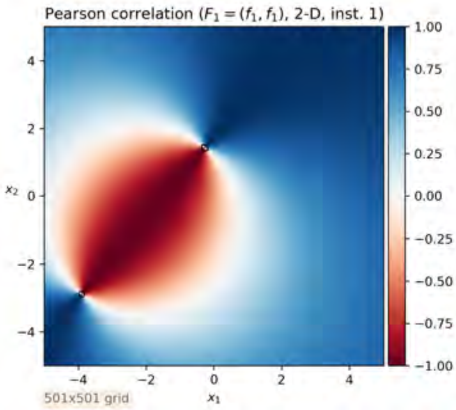



**Figure 14** The grid point $x = (x_1, x_2)$ and the $p$ points (shown in orange) used in the computation of the Pearson correlation coefficient (here, $p = 15$).

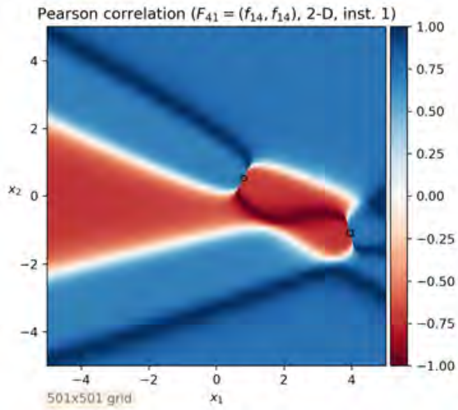
#### 4.8.3.2 Problems with two objectives

First, we wish to explore the simplest case of two objectives. For this, we select six bi-objective problems from the `bbob-biobj` suite of benchmark problems [2]. They are constructed by combining two single-objective functions from the `bbob` suite [8]. Figure 15 shows the visualization of correlations between the two objectives for each of the six problems.

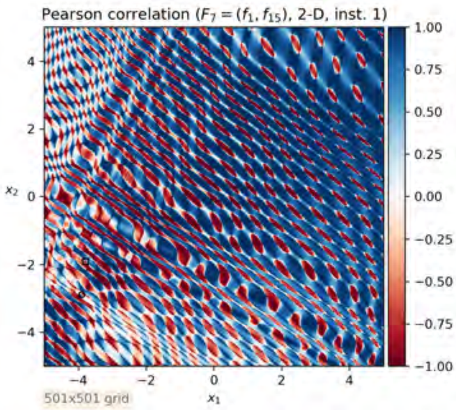
The double sphere problem $F_1 = (f_1, f_1)$, where $f_1$ is the `bbob` sphere function, is a unimodal problem with a known Pareto set – the line segment connecting the two single-objective optima. We can see from the correlation plot in Figure 15a the expected outcome – close to the Pareto set, the objectives are anti-correlated (red hues), while further away they are correlated (blue hues).



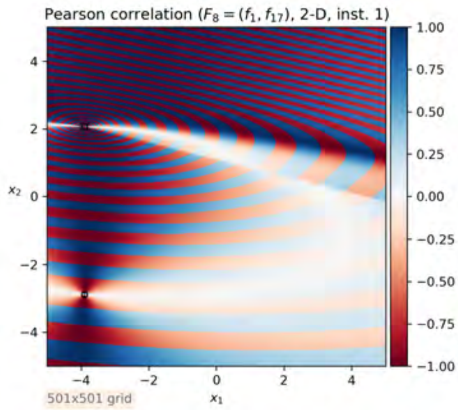

**(a)** The double sphere problem $F_1$.



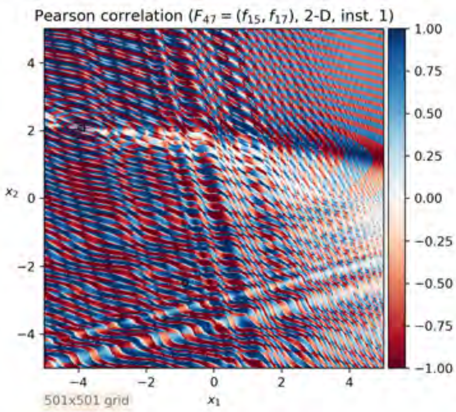

**(b)** The double different powers problem $F_{41}$.



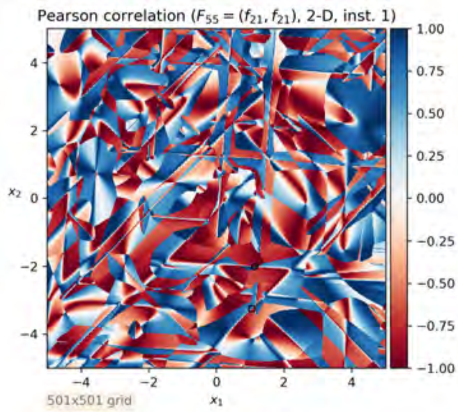

**(c)** The sphere-Rastrigin problem $F_7$.




**(d)** The sphere-Schaffer problem $F_8$.




**(e)** The Rastrigin-Schaffer problem $F_{47}$.




**(f)** The double Gallagher problem $F_{55}$.

**Figure 15** Person correlation coefficient for some chosen 2-D `bbob-biobj` problems (these and other plots for `bbob-biobj` problems will be made available at `https://numbbo.github.io/bbob-biobj/vis/`). Blue hues denote positive correlations, red hues negative ones and white indicates no correlation.

In the problem $F_{41} = (f_{14}, f_{14})$, both objectives are unimodal as well, but they correspond to the bbob sum of different powers function $f_{14}$, which is non-separable and ill-conditioned. Figure 15b shows that in this case, the objectives are anti-correlated also far away from the Pareto set.

The next two problems are a combination of a unimodal objective (the bbob sphere function $f_1$) and a highly multimodal one. In the problem $F_7 = (f_1, f_{15})$, this is the bbob Rastrigin function $f_{15}$, while in the problem $F_8 = (f_1, f_{17})$, it is the bbob Schaffer F7 function $f_{17}$ with condition number 10. In both instances, visualized in Figures 15c and 15d, the resulting bi-objective problems have multiple disconnected regions of the search space where the objectives are anti-correlated.

Finally, in the last two selected problems, both objectives are highly multimodal. The problem $F_{47} = (f_{15}, f_{17})$ combines the bbob Rastrigin function $f_{15}$ with the bbob Schaffer F7 function $f_{17}$ with condition number 10 and the problem $F_{55} = (f_{21}, f_{21})$ two bbob Gallagher's Gaussian functions $f_{21}$ with 101 median peaks. We can see from the correlation plots in Figures 15e and 15f the high number of disconnected regions of anti-correlated objectives.

These examples challenge some of our preexisting notions about the correlation between objectives. In particular, they show that it is closely connected to the problem multimodality – understandably, given that the correlation between two objectives equals $-1$ at any locally optimal set. In fact, the notion of a globally (i.e., Pareto) optimal set is inconsequential for correlation values. It is therefore rather meaningless to discuss correlations between objectives without taking into account their multimodality. We also see that the Pearson correlation coefficient values are themselves positively correlated with the length of the normalized bi-objective gradient as defined in [10] and visualized in [2].

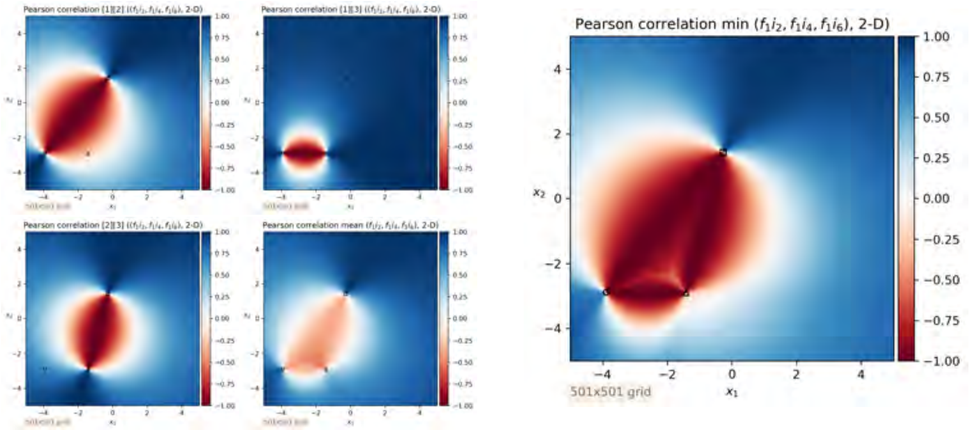
#### 4.8.3.3 Problems with three objectives

The Pearson correlation coefficient is defined only for two objectives. When the objectives are three (or more), we can compute all their pairwise correlations. We wish to visualize their minimal values to emphasize parts of the search space with the highest anti-correlation as they are locally optimal.

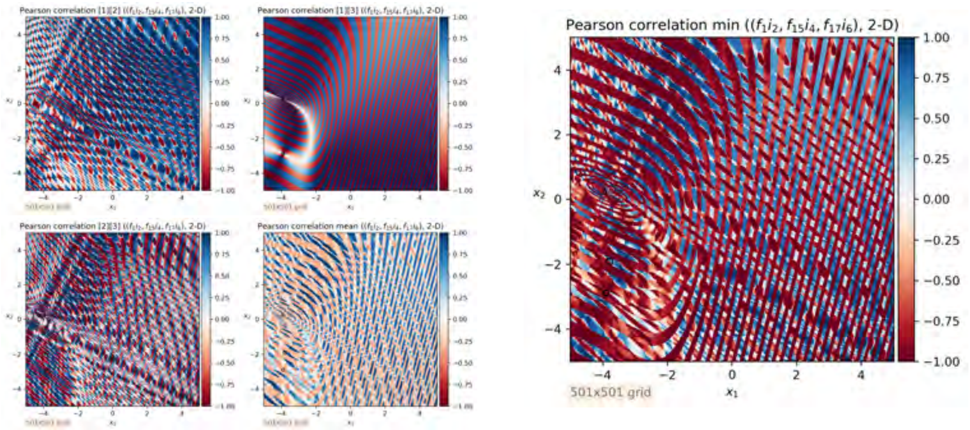
Exemplary three-objective problems are again constructed by combining bbob functions – now three. This time, we chose the triple sphere problem, the sphere-Rastrigin-Schaffer problem and the triple Gallagher problem. See Figure 16 for their visualizations. For each problem we show on the left hand side the pairwise correlations for objectives 1 and 2, objectives 1 and 3 and objectives 2 and 3 as well as their mean. On the right hand side, their minimum is presented.

The Pareto set of the triple sphere problem is the triangle spanned by the three single-objective optima. From Figure 16a we see that its minimal pairwise Pearson correlation coefficient equals $-1$ only at the edges of this triangle, not in its interior. This shows that, unlike in the bi-objective case, one cannot rely on pairwise correlations alone to infer local optimality of a solution in case of more than two objectives. A procedure similar to the one from [12] should be tried to amend this issue.

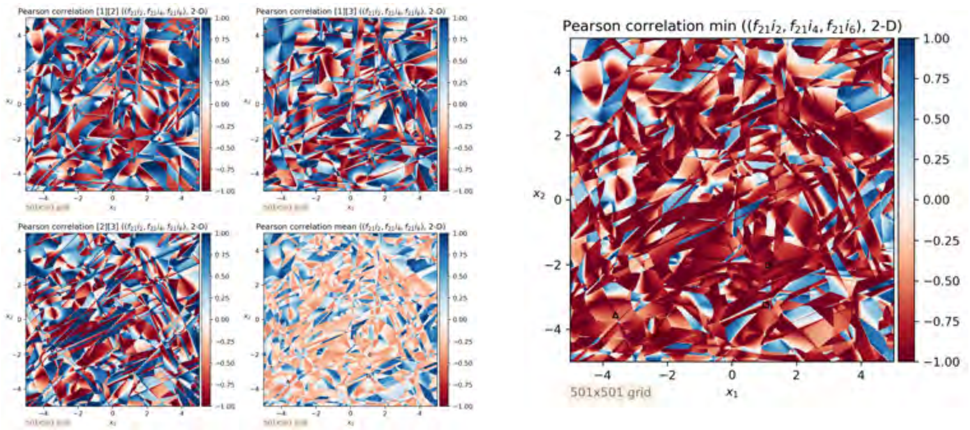
Further examples show the minimal pairwise correlation for the sphere-Rastrigin-Schaffer problem (Figure 16b) and the triple Gallagher problem (Figure 16c). Both are highly multimodal, resulting in many disconnected regions with anti-correlated pairs of objectives (red hues).




**(a)** The triple sphere problem.




**(b)** The sphere-Rastrigin-Schaffer problem.




**(c)** The triple Gallagher problem.

**Figure 16** Visualization of correlations for three three-objective problems. Smaller plots from top to bottom, left to right: pairwise Pearson correlation coefficients for objectives 1 and 2, 1 and 3 and 2 and 3, and their mean. Larger plot: minimum value of the pairwise correlation coefficients.

#### 4.8.3.4 Problems with five objectives

We next consider a couple of planar problems with five objectives, using the distance-based multi-objective point problem (DBMOPP) generator [5]. This generator allows us to create problem instances which natively live in 2-D (or map to 2-D), which can have an arbitrary number of objectives and can exhibit a range of other problem properties.

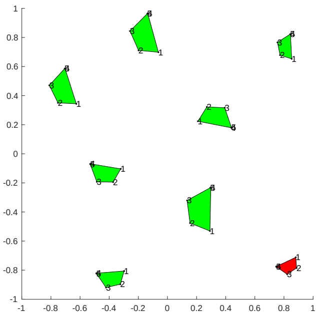
We first generate a box-constrained instance with a single spatially contiguous Pareto set (shown in red in Figure 17a) and seven other regions which generate local fronts of the same shape, but which are dominated (shown in green in Figure 17a). Figure 17b shows the corresponding *dominance landscape* [4]. Black regions in this figure show locations which are not dominated by any immediate neighbor (dominance neutral regions). Gray regions in contrast denote locations which have at least one dominating neighbor, but where all point-based dominance hill-climbs (by moving to an adjacent dominating neighbor) lead to the same dominance neutral region – different shades of gray are used to distinguish these different *basins*. White regions signify where point-based dominance hill-climbs lead to multiple different dominance neutral regions (effectively multi-objective saddle-points), depending on which chain of dominating neighbors one follows. Figure 17c shows the dominance ratio [6] landscape for the problem instance. In this plot, the value at a location denotes the proportion of the entire domain which *weakly dominates* it (i.e. dominates or is equal to it). That is, a value of 0.0 will indicate a location is Pareto optimal, whereas a value of 0.2 indicates that 20% of the domain relates to locations with equal or better performance on all criteria. Pearson correlation plots are shown in Figures 17d–17f. For this problem we can see the eight distinct local optima regions clearly in the Dominance ratio plot, with the induced dominance neutral plateaus between these regions additionally identifiable in the dominance landscape and correlation plots.

The second example shown in Figure 18a has a single spatially contiguous Pareto set region (red), 3 dominance resistance regions (blue), 3 local fronts regions (green) and 30% of the decision space is designed as being flat under the objectives (cyan). The corresponding dominance landscape is shown in 18b, and the dominance ratio landscape in 18c. Pearson correlation plots are shown in Figures 18d–18f. The impact of the flat objective regions is clear across the plots, and all views of the landscape are considerably more cluttered due to the interactions of the various problem features.

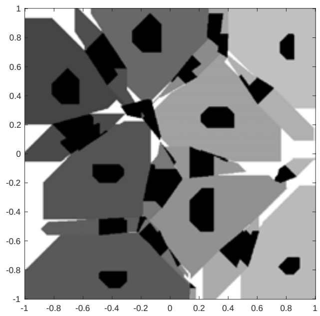
### 4.8.4 Conclusions

We recognized that evaluation times can differ among solutions of expensive real-world problems. We were therefore interested in exploring whether the correlation between objectives and their evaluation times can be used to save time-consuming function evaluations. A deeper look into the properties of some real-world applications has shown that a general model for such a correlation is hard to find. Therefore, the evaluation time was regarded as an additional objective to be minimized.

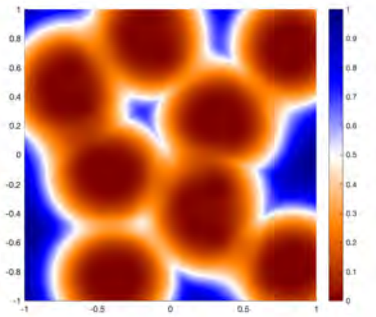
Next, we researched the correlation between objectives, estimating it with the Pearson correlation coefficient. To gain a better understanding of the distribution of its values in the search space, we visualized them for a number of test problems with two variables and two, three and five objectives. The visualizations have shown that some of our intuition about the correlation between objectives was wrong. For example, we could find unimodal problems with anti-correlated objectives not only close to the Pareto set, but also far away from it. Visualizations of multimodal problems have proven that many distinct anti-correlated regions can be located throughout the search space, surrounded by regions with correlated



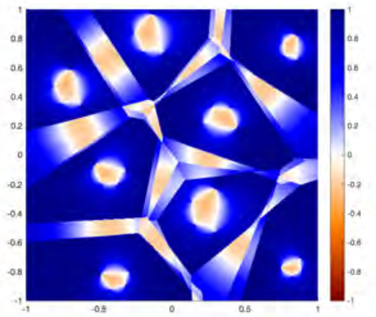
**(a)** Problem configuration.



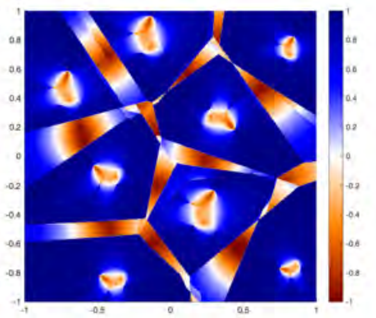
**(b)** Dominance landscape.



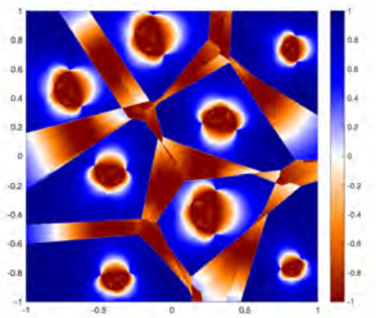
**(c)** Dominance ratio.



**(d)** Mean Pearson coefficient.

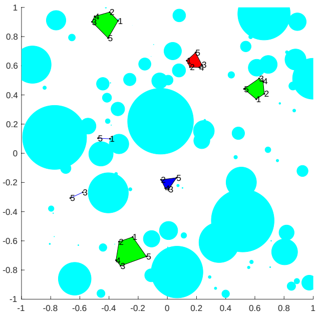


**(e)** Median Pearson coefficient.

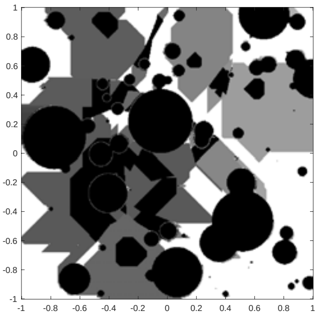


**(f)** Min Pearson coefficient.

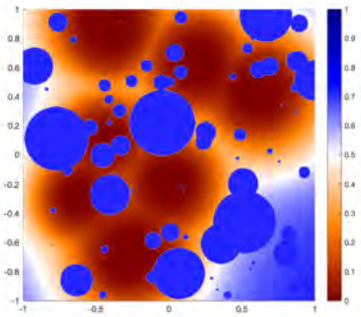
**Figure 17** Problem plots and Person correlation values for a 5-objective 2-D `DBMOPP` instance. In the correlation plots blue hues denote positive correlations, red hues negative ones and white indicates no correlation.



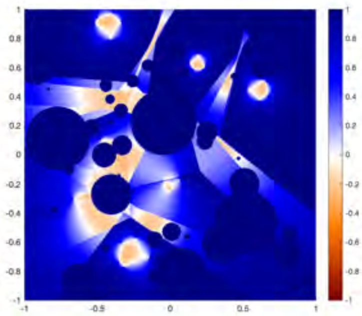
**(a)** Problem configuration.



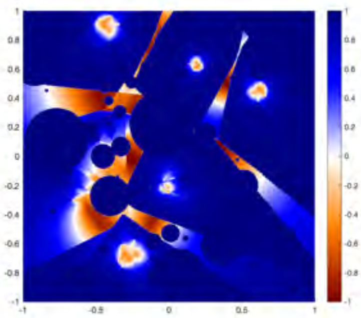
**(b)** Dominance landscape.



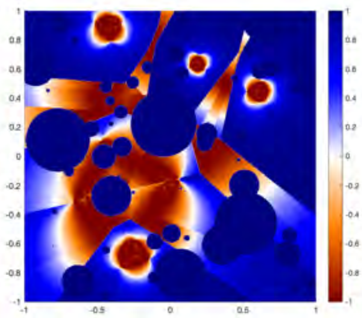
**(c)** Dominance ratio.



**(d)** Mean Pearson coefficient.



**(e)** Median Pearson coefficient.



**(f)** Min Pearson coefficient.

**Figure 18** Problem plots and Person correlation values for a more complex 5-objective 2-D DBMOPP instance. In the correlation plots blue hues denote positive correlations, red hues negative ones and white indicates no correlation.

objectives. In fact, the visualizations have demonstrated that correlation is closely tied to the problem multimodality and has a nonlinear monotonous relation with the length of the bi-objective gradient. Finally, while pairwise anti-correlations between objectives correspond to the locally optimal solutions for problems with two objectives, this is no longer the case when the number of objectives is three or more.

## References


1 Richard Allmendinger and Joshua D. Knowles. Heterogeneous objectives: State-of-the-art and future research. *CoRR*, abs/2103.15546, 2021.

2 Dimo Brockhoff, Anne Auger, Nikolaus Hansen, and Tea Tušar. Using well-understood single-objective functions in multiobjective black-box optimization test suites. *Evolutionary Computation*, 30(2):165–193, 2022.

3 Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. Neural architecture search: A survey. *Journal of Machine Learning Research*, 20(55):1–21, 2019.

4 Jonathan E. Fieldsend, Tinkle Chugh, Richard Allmendinger, and Kaisa Miettinen. A feature rich distance-based many-objective visualisable test problem generator. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '19, page 541–549, New York, NY, USA, 2019. Association for Computing Machinery.

5 Jonathan E. Fieldsend, Tinkle Chugh, Richard Allmendinger, and Kaisa Miettinen. A visualizable test problem generator for many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 26(1):1–11, 2022.

6 Carlos Manuel Mira da Fonseca. *Multiobjective genetic algorithms with application to control engineering problems.* PhD thesis, University of Sheffield, 1995.

7 Salvatore Greco, Kathrin Klamroth, Joshua D. Knowles, and Günter Rudolph. Understanding complexity in multiobjective optimization (Dagstuhl Seminar 15031). *Dagstuhl Reports*, 5(1):96–163, 2015.

8 Nikolaus Hansen, Steffen Finck, Raymond Ros, and Anne Auger. Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Research Report RR-6829, INRIA, 2009.

9 Miha Hafner, David Rajšter, Marko Žibert, Tea Tušar, Bernard Ženko, Martin Žnidaršič, Flavio Fuart, and Daniel Vladušič. Artificial intelligence support for tunnel design in urban areas. In *Tunnels and Underground Cities: Engineering and Innovation meet Archaeology, Architecture and Art- Proceedings of the WTC 2019 ITA-AITES World Tunnel Congress*, pages 2196–2205, 2019.

10 Pascal Kerschke and Christian Grimme. An expedition to multimodal multi-objective optimization landscapes. In *Evolutionary Multi-Criterion Optimization – 9th International Conference, EMO 2017, Münster, Germany, March 19-22, 2017, Proceedings*, volume 10173 of *Lecture Notes in Computer Science*, pages 329–343. Springer, 2017.

11 Boris Naujoks, Werner Haase, Jörg Ziegenhirt, and Thomas Bäck. Multi objective airfoil design using single parent populations. In *GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference, New York, USA, 9-13 July 2002*, pages 1156–1163. Morgan Kaufmann, 2002.

12 Lennart Schäpermeier, Christian Grimme, and Pascal Kerschke. To boldly show what no one has seen before: A dashboard for visualizing multi-objective landscapes. In *Evolutionary Multi-Criterion Optimization – 11th International Conference, EMO 2021, Shenzhen, China, March 28-31, 2021, Proceedings*, volume 12654 of *Lecture Notes in Computer Science*, pages 632–644. Springer, 2021.

13 Tea Tušar, Peter Korošec, and Bogdan Filipič. A multi-step evaluation process in electric motor design. In *Slovenian Conference on Artificial Intelligence, Proceedings of the 26rd International Multiconference Information Society – IS 2023*, volume A, pages 48–51. Jožef Stefan Institute, 2023.

14 Vanessa Volz, Boris Naujoks, Pascal Kerschke, and Tea Tušar. Single- and multi-objective game-benchmark for evolutionary algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2019, Prague, Czech Republic, July 13-17, 2019*, pages 647–655. ACM, 2019.

## 5 Seminar Schedule

### Monday, September 4, 2023

**09:00–10:30** Welcome Session

- Welcome and introduction
- Short presentation of all participants

Coffee Break

**11:00–12:00** Seminar Overview

- Seminar scope
- Offers-and-needs market

Lunch

**14:00–15:30** Modelling to Address Budget Constraints

- Juliane Mueller: *Surrogate model guided optimization of expensive black-box multiobjective problems*
- Andrea Raith: *Problem decomposition in biobjective optimization*

Coffee Break

**16:00–16:30** Working Group Formation

**16:30–18:00** Working Groups

Dinner

### Tuesday, September 5, 2023

**09:00–09:30** Reporting from Small Working Groups Chair: Matthias Ehrgott

**09:30–10:30** Small Working Groups

Coffee Break

**11:00–12:00** Heuristic Optimization and Human Involvement Chair: Jürgen Branke

- Thomas Bäck: *How to help end users when the budget is limited?*
- Robin Purshouse: *Towards decision analytic workflows for real-world problems: Simulation model calibration and multi-objective optimization on a shared evaluation budget*
- Benjamin Doerr: *Runtime analysis for the NSGA-II*
- Kaisa Miettinen: *Perspectives to dealing with computationally expensive multiobjective optimization problems*

Lunch

**14:00–15:30** Small Working Groups

Coffee Break

**16:00–17:00** Small Working Groups

**17:00–18:00** Reporting from Small Working Groups and General Discussion
Dinner

### Wednesday, September 6, 2023

**09:00–10:30** Small Working Groups
Coffee Break

**11:00–12:00** Approximation and Exact Methods Chair: Karl-Heinz Küfer

- Kathrin Klamroth: *Objective space methods: Pareto front approximations on a budget*
- Michael Stiglmayr: *Multi-objective branch-and-bound on a budget*
- Frank Neumann: *Fast Pareto optimization using sliding window selection*
- Alma Rahat: *Efficient approximation of expected hypervolume improvement using Gauss-Hermite quadrature*

**12:00-12:15** Group Photo Outside
Lunch

**13:30–15:30** Hiking Trip
Coffee Break

**16:00–18:00** Small Working Groups
Dinner

### Thursday, September 7, 2023

**09:00–10:00** Reporting from Small Working Groups Chair: Kaisa Miettinen
Coffee Break

**10:30–12:00** Small Working Groups
Lunch

**14:00–15:30** Small Working Groups
Coffee Break

**16:00–16:30** Graphical and probabilistic approaches Chair: Boris Naujoks

- Ralph Steuer: *A visualization-aided approach for solving tri-criterion portfolio problems*
- Hao Wang and Kaifeng Yang: *Probability of "improvement" in multi-objective Bayesian optimization*

**16:30–17:00** Announcements

**17:00–18:00** Joint session with DS22362
Dinner

**20:00–23:00** Informal Get Together (BYOB, meet in the cafeteria)

### Friday, September 8, 2023

**09:00–10:30** Final Reporting from Working Groups
Coffee Break

**11:00–12:00** Closing Session

## 6 Topics of interest for participants for next Dagstuhl Seminar

In the closing session on Friday, the participants reflected upon their experience and presented their ideas on a potential future seminar that would leverage the progress made during the current one. During this discussion, some topics appeared to center around "Artificial Intelligence (AI)". A two-way perspective was suggested: AI for multiobjective optimization and multiobjective optimization for AI. Another suggestion was to focus on the "gap" between the industrial and the academic practice of multiobjective optimization. This suggestion was well-received by both industrial and academic participants of the seminar as the focus during the week was on a "budget" that might also mean decision maker's limitations. Focusing on how the theoretical and methodological achievements on the academic front can be made more accessible to practitioners in industry may be a future direction to pursue. This direction will also possibly require placing more emphasis on modelling, handling the noise, errors and uncertainties in the process. The organizers will use these suggestions as the basis for their discussion about possible topics for the next edition of this seminar series and for the preparation of a proposal for a continuation of the series.

## 7 Changes in the seminar organization body

As part of a continuing effort to renew the organizing board of this series of Dagstuhl Seminars, Margaret Wiecek steps down from the team of organizers, a role that she has held for three terms of office. On behalf of all the participants of the seminar, Richard Allmendinger, Carlos Fonseca and Serpil Sayin would like to express appreciation to Margaret for her contributions and leadership that have been fundamental for the series success.

We are pleased to announce that our esteemed colleague and a multiple-times Dagstuhl attendee Susan Hunter has agreed to serve as a co-organizer for future editions of this Dagstuhl Seminar series on Multiobjective Optimization. We look forward to collaborating with her in the near future.

## Participants

- Thomas Bäck
Leiden University, NL
- Mickaël Binois
INRIA – Sophia Antipolis, FR
- Fritz Bökler
Universität Osnabrück, DE
- Jürgen Branke
University of Warwick, GB
- Dimo Brockhoff
INRIA Saclay – Palaiseau, FR
- Tinkle Chugh
University of Exeter, GB
- Kerstin Dächert
HTW Dresden, DE
- Benjamin Doerr
Ecole Polytechnique – Palaiseau, FR
- Matthias Ehrgott
Lancaster University, GB
- Gabriele Eichfelder
TU Ilmenau, DE
- Jonathan Fieldsend
University of Exeter, GB
- Carlos M. Fonseca
University of Coimbra, PT
- Susan R. Hunter
Purdue University, US
- Ekhine Irurozki
Telecom Paris, FR
- Hisao Ishibuchi
Southern Univ. of Science and Technology – Shenzen, CN
- Andrzej Jaszkiewicz
Poznan University of Technology, PL
- Pascal Kerschke
TU Dresden, DE
- Kathrin Klamroth
Universität Wuppertal, DE
- Joshua D. Knowles
Schlumberger Cambridge Research, GB
- Karl Heinz Küfer
Fraunhofer ITWM – Kaiserslautern, DE
- Arnaud Liefooghe
University of Lille, FR
- Kaisa Miettinen
University of Jyväskylä, FI
- Juliane Mueller
NREL – Golden, US
- Boris Naujoks
TH Köln, DE
- Aneta Neumann
University of Adelaide, AU
- Frank Neumann
University of Adelaide, AU
- Markus Olhofer
HONDA Research Institute Europe GmbH – Offenbach, DE
- Luís Paquete
University of Coimbra, PT
- Robin Purshouse
University of Sheffield, GB
- Alma Rahat
Swansea University, GB
- Andrea Raith
University of Auckland, NZ
- Enrico Rigoni
ESTECO SpA – Trieste, IT
- Stefan Ruzika
RPTU – Kaiserslautern, DE
- Serpil Sayin
Koc University – Istanbul, TR
- Anita Schöbel
Fraunhofer ITWM – Kaiserslautern, DE
- Britta Schulze
Universität Wuppertal, DE
- Ralph E. Steuer
University of Georgia, US
- Michael Stiglmayr
Universität Wuppertal, DE
- Tea Tusar
Jozef Stefan Institute – Ljubljana, SI
- Daniel Vanderpooten
University Paris-Dauphine, FR
- Vanessa Volz
modl.ai – Copenhagen, DK
- Hao Wang
Leiden University, NL
- Margaret M. Wiecek
Clemson University, US
- Kaifeng Yang
Univ. of Applied Sciences – Hagenberg, AT

Dagstuhl Seminar 23362

# Decision-Making Techniques for Smart Semiconductor Manufacturing

**Hans Ehm**[*1], **John Fowler**[*2], **Lars Mönch**[*3], **and Daniel Schorn**[†4]

1 **Infineon Technologies – München, DE.** hans.ehm@infineon.com
2 **Arizona State University – Tempe, US.** john.fowler@asu.edu
3 **FernUniversität in Hagen, DE.** lars.moench@fernuni-hagen.de
4 **FernUniversität in Hagen, DE.** daniel-sascha.schorn@fernuni-hagen.de

**Abstract**

In September 2023 the Dagstuhl Seminar 23362 explored the needs of the semiconductor industry for novel decision-making techniques and the related information systems to empower flexible decisions for smart production. The seminar participants also spent time identifying requirements for a simulation testbed which allows for assessing smart planning and control decisions in the semiconductor industry. The Executive Summary describes the process of the seminar and discusses key findings and areas for future research regarding these topics. Abstracts of presentations given during the seminar and the output of breakout sessions are collected in further sections.

**Seminar** September 3–8, 2023 – https://www.dagstuhl.de/23362

**2012 ACM Subject Classification** Applied computing → Industry and manufacturing; Computing methodologies → Modeling and simulation; Applied computing → Multi-criterion optimization and decision-making

**Keywords and phrases** analytics, modeling, semiconductor manufacturing, simulation, smart manufacturing

**Digital Object Identifier** 10.4230/DagRep.13.9.69

## 1 Executive Summary

*Hans Ehm*
*John Fowle*
*Lars Mönch*



The Industry 4.0 vision is a frequently discussed topic in manufacturing enterprises in Europe, Asia, and North America. The Industry 4.0 vision is a frequently discussed topic in manufacturing enterprises in Europe, Asia, and North America. It is expected that advanced technologies such as Cyber-Physical Systems, Internet of Things, cloud computing, and big data technologies enable the emergence of smart manufacturing systems. A smart factory promises to bring transparency to manufacturing facilities by integrating technological advances in computer networks, data integration, and analytics. At the same time, critical questions are asked related to the benefits of Industry 4.0. It is mainly criticized that the requirements and consequences of Industry 4.0 regarding future production planning and control strategies are not fully understood or not even taken into account in the overall Industry 4.0 conception, i.e., many of key decision processes are not included.

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


Decision-Making Techniques for Smart Semiconductor Manufacturing, *Dagstuhl Reports*, Vol. 13, Issue 9, pp. 69–102
Editors: Hans Ehm, John Fowler, and Lars Mönch
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The semiconductor industry is capital intensive. The manufacturing process is very complex due to reentrant flows in combination with very long cycle times and multiple sources of uncertainty. This industry is an extreme field for production planning and control solutions from an algorithmic point of view, as well as from a software and information systems point of view. The degree of automation was always – and is still – high compared to other industries. On the one hand, one can argue that in wafer fabs elements of smart manufacturing are already realized, namely most of manufacturing information is available in real-time, the manufacturing process is paperless, lots can be uniquely identified and located, and collaborative human-machine interaction exists. On the other hand, there are significant differences in automation efforts related to manual work-intensive industries such as automotive or aircraft manufacturing where assembly operations are performed in flow lines. In addition to shop-floor control concerns, supply chain management problems have become more and more important which necessitate a horizontal integration of the semiconductor supply chain and digital transformation for the industry ecosystem.

The major objective of this Dagstuhl Seminar was related to developing a research agenda for making smart semiconductor manufacturing decisions and the information systems to empower flexible decisions for smart production. The research agenda was developed around the following two main topics:

**Topic 1:** Novel decision-making approaches that exploit the huge amount of available data and orchestrate the interrelated decisions

**Topic 2:** Future information systems for decision support and facilitating digital transformation.

The purpose of this seminar was to bring together researchers from different disciplines including information systems, computer science, industrial engineering, supply chain management, data science, and operations research whose central interest is in decision-making for smart semiconductor manufacturing. Moreover, practitioners from the semiconductor industry who have frequently articulated their perception that academic research did not always address the real problems faced by the industry brought in their domain knowledge to make sure that progress towards applicability and feasibility was made during this seminar. Detailed introduction to the topic, the objectives, and results of the seminar, as well as the next steps will be presented in the following sections of this report.

## 2 Table of Contents

## 3 Seminar Objectives

*Hans Ehm*
*John Fowle*
*Lars Mönch*


Complex manufacturing processes are the heart of semiconductor manufacturing. A semiconductor chip is a highly miniaturized, integrated circuit (IC) consisting of thousands of components. Semiconductor manufacturing starts with thin discs, called wafers, made of silicon. A large number of usually identical chips can be produced on each wafer by fabricating the ICs layer by layer in a wafer fabrication facility (wafer fab). The corresponding step is referred to as the Fab step. Next, electrical tests that identify the individual dies that are likely to fail when packaged are performed in the Probe facility. An electronic map of the condition of each die is made so that only the good ones will be used. The probed wafers are then sent to an Assembly facility where the good dies are put into an appropriate package. The assembled dies are sent to a test facility where they are tested to ensure that only good products are sent to customers. Wafer fabrication and probe are often called the front-end and assembly and test are called the back-end.

The semiconductor industry is capital intensive caused primarily by extremely expensive machines, some up to \$ 100 million US. The manufacturing process is very complex due to the reentrant flows in combination with very long cycle times and the multiple sources of uncertainty involved. The demand is highly volatile. The consequences of the current chip shortage are felt throughout the economy and in everyday life. The semiconductor industry is an extreme field for decision support solutions from an algorithmic as well as from a software and information systems point of view. The huge size of the supply chains involved, the pervasive presence of different kinds of uncertainties, and the rapid pace of change leads to an environment that places approaches developed in other industries under major stress. Modeling and analysis approaches that are successful in this industry are likely to find applications in other areas, and to significantly advance the state of the art in their fields (Chien et al. 2011).

The principle architecture of the planning and control system of a conventional semiconductor supply chain consisting of several wafer fabs/sort facilities and assembly and test (A&T) facilities is shown in Figure 1.

The first objective of the seminar consisted of developing a research agenda for decision-making in smart semiconductor manufacturing. This included innovative modeling approaches for supply chain planning and detailed production planning and scheduling/dispatching in semiconductor supply chains. But it also included ideas on how to design the related future information systems. Proposing such a research agenda is timely, since on the one hand there is a technology pressure and on the other hand a demand pull for advanced decision-making procedures that support digitalization efforts in semiconductor manufacturing.

The proposed research agenda is not only important for semiconductor manufacturing, because decision-making approaches that are successful in this industry are likely to find applications in other areas. We expect that these applications will significantly advance the state of the art in their fields.

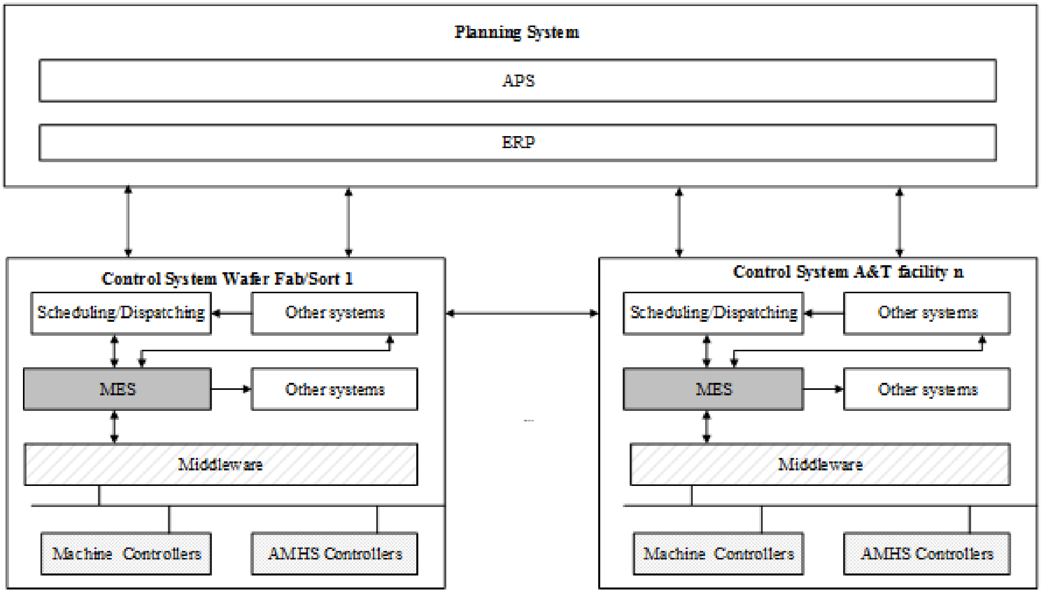



**Figure 1** Planning and Control System of a Semiconductor Supply Chain (adapted from Mönch et al. 2013).

The developed research agenda is around the following two main topics:

- **Topic 1: Novel decision-making approaches that exploit the huge amount of available data and orchestrate the interrelated decisions:**
  - Which parts of the Industry 4.0 and/or the smart manufacturing vision are already implemented in semiconductor manufacturing and what is still missing?
  - What are the specific automation drivers in semiconductor manufacturing compared to other industries?
  - Can techniques from smart manufacturing help to reduce the current chip shortage?
  - Which additional data, for instance, provided by sensors and cyber-physical systems can be used to make better decisions (Chien and Chuang 2014)? How can the improvement potential based on the advanced data availability be quantified (Khakifirooz et al. 2018)?
  - Which decisions can or even should be integrated? Possible examples for integrated short-term decisions are job scheduling on machines and automated transportation and job scheduling and preventative maintenance planning. Integrated scheduling and process control decisions are another example. Energy-aware scheduling approaches require integrated decisions too (Rocholl et al. 2020, Rocholl and Mönch 2021). On the mid-term planning level, the integrated management of production jobs and engineering jobs is challenging. Up to 30% of all jobs in a wafer fab are engineering jobs. They com¬pete with the production jobs for the scarce capacity of the machines. It is also interesting to make integrated production planning and inventory planning decisions in semiconductor manufacturing.
  - Which changes are required or are even possible for planning and control algorithms in smart manufacturing systems? Do we expect fundamentally new algorithms?
  - Is there a need for new fab layouts in the context of smart manufacturing? Initial steps towards the possible redesign of the automated material handling system (AMHS) are discussed by Ham and Kim (2017) and Hwang and Jang (2020).

  - How can dynamics and stochasticity be included into decision-making? Different ways to anticipate stochasticity including robust optimization, approximate dynamic programming, and stochastic programming have to be researched in the smart semiconductor manufacturing context. Different ways to appropriately deal with stochasticity including rolling planning techniques and inventory holding strategies have to be studied. Generation of scenarios and other distribution parameters for planning problems in supply chains using big data techniques have to be researched.
  - Many planning and control approaches are based on (distributed) hierarchical approaches. What is the role of anticipation of lower level behavior in upper level decision-making? Because many different, often autonomous decision-making entities including humans occur in semiconductor manufacturing, negotiation approaches are typical in such distributed hierarchical planning and control systems How can such negotiation approaches be automated and which decisions should continue to be made by humans?
  - How can sustainability issues be incorporated into decision making? For instance, taking advantage of real-time pricing in future energy markets is only reasonable when scheduling decisions can be made in real-time.
  - What is the relationship of real-time decisions based on real-time information on the status of the shop-floor (or even the supply chain) and planning nervousness?
  - As the level of automation increases in the factory of the future, there is a need to adapt the decision-making entities to the current situation at the shop floor and the entire supply chain. Which machine learning paradigms are appropriate to reach this goal (Chien et al. 2021)?

- **Topic 2: Future information systems for decision support and facilitating digital transformation:**
  - What changes for next-generation decision support systems are required? It is expected that decentralized decision support systems are more important than in the past.
  - Can advanced information systems help to reduce the current chip shortage?
  - What alternative software solutions including software agents and service-oriented computing for planning and scheduling applications in smart semiconductor manufacturing are beneficial?
  - What is the role of different simulation paradigms in the factory of the future/supply chain of the future?
  - What is the expected benefit of digital twins in semiconductor manufacturing? For instance, it has to be decided at what levels (e.g. factory, supply chain) they should be considered.
  - What integration concepts for state-of-the-art computing techniques to obtain models that are computationally tractable and address the different uncertainties encountered in this industry are appropriate for their usage in smart semiconductor manufacturing?
  - What interaction of human agents with information systems in the factory of the future is beneficial?
  - Because of the complexity of semiconductor supply chains, long computing times still hinder the usage of analytic solution approaches especially for what-if analysis. What is the role of state-of-the-art computing techniques including parallel computing on Graphics Processing Units (GPU) machines or cloud computing techniques in decision-making for smart semiconductor manufacturing?

Since the expected potential of smart manufacturing is based on advanced information and communication technologies, we thought that the second topic is important and should be also addressed in the research agenda. Research related only to the first main topic is not sufficient since it is expected that technologies such as cyber-physical systems, software agents, cloud computing, and simulation are technological enablers for the novel decision-making paradigms of the first topic.

Due to the inherent complexity of semiconductor supply chains it requires simulation of the physical supply chain to understand the interactions between the planning and control components and the physical supply chain, to find solution approaches to problems, and to verify them in the risk-free simulation environment before implementing them. There are widely accepted reference (simulation) models for single wafer fabs and simple semiconductor supply chains. These models are primarily based on simulation models proposed in the Measurement and Improvement of Manufacturing Capacity (MIMAC) project (led by one of the organizers of this proposed Dagstuhl Seminar) 25 years ago that are still used by many academic researchers working with the semiconductor industry. These models do not reflect the complexity and the level of detail of current and future semiconductor supply chains. Even more recent simulation models such that the models from the SMT 2020 testbed (Kopp et al. 2020) are not fully appropriate for smart manufacturing since they do not support, for instance, AMHS- or sustainability-related decisions. Therefore, the second goal of the seminar consisted in identifying the core elements of a simulation testbed which allows for assessing smart planning and control decisions in the semiconductor industry.
The second objective can be reached by the following steps:

- specification of the main ingredients of the simulation models of the testbed ,
- specification of additional requirements compared to conventional wafer fab or semiconductor supply chain simulation models that arise from the smart semiconductor manufacturing context, for instance, providing data gathering schemes that allow to mimic the application of big data analytics,
- design of rich reference application scenarios.

Of course, this work was not completed during the seminar, but because of the working groups we were able to come up with a significant draft that can be refined in various ways after the seminar. We believe that the research agenda is what is important and that the simulation testbed model for assessing smart decision-making procedures is simply a means to that end.

## 3.1 The Process

In the opening session, the organizers welcomed the participants. Next, the participants each introduced themselves. An overview of the goals and objectives of the seminar and a detailed review of the seminar program including the ground rules for interactions followed after the introduction. The remainder of the day on Monday consisted of an introduction into smart manufacturing (by John Fowler and Lars Mönch) and six industry overview talks (by Adar Kalir, Marcel Stehli, Alexandru Prisacaru, Thomas Ponsignon, Hans Ehm, and Peter Lendermann). Tuesday and half a day on Wednesday were devoted to presentations and discussions about the various elements of the semiconductor supply chain planning and control systems shown in Figure 1 above and their relations to smart semiconductor manufacturing. See Table 1 below for a list of topics and presenters and Appendix B for abstracts of the presentations.

**Table 1** Individual Presentations.

| Topic | Presenter |
|---|---|
| Machine Learning-based Process Modeling in Semiconductor Manufacturing | Gian Antonio Susto |
| Demonstration of the Feasibility of the Application of Machine Learning for Production Scheduling | Cathal Heavey |
| Advancing Automation through Robot Collaboration Intelligence and Digital Twin Integration | Young Jae Jang |
| Digital Twins | Andrea Matta |
| SMT2020 Reference Model | Michael Hassoun |
| Cloud-based Simulation Experiments for Optimization and Machine Learning | Oliver Rose |
| Order-lot Pegging in a Multi-fab Setting | Liji Shen |
| Agent-based Decision Support in Borderless Fab Scenarios in Semiconductor Manufacturing | Raphael Herding |
| DTFab: Performance Improvement, Analytics and Security in DT-controlled Semiconductor Systems | Giulia Pedrielli |
| Complex flexible Job-shop Scheduling Problems and Semiconductor Manufacturing | Stephane Dauzère-Pérès |
| Scheduling in Semiconductor Reliability Testing Labs | Jessica Hautz |
| Minimizing Makespan for a Multiple Orders per Job Scheduling Problem in a Two-stage Permutation Flow Shop | Rohan Korde |
| Genetic Programming for Energy-aware Scheduling | Daniel Schorn |
| Reinforcement Learning | Mahsa Shekari |

Wednesday afternoon was the excursion to Trier that was enjoyed by the participants.

Thursday was devoted to a set of three breakout sessions with report outs on the topics in Table 2. Appendix C has the breakout report outs.

The first set of breakout sessions had three groups focus on machine learning (ML) since ML techniques are an important element of smart manufacturing. The second set of breakouts had again three groups consider simulation-based decision support since such techniques are an important element of smart manufacturing. The third breakout session had again three groups, one dealing with sustainability issues as a core element of smart manufacturing, one with information systems including ontologies, and a last one that prepared a joint session with the participants of the Dagstuhl Seminar running in parallel with the present seminar. During the joint session with the other seminar on Thursday evening we learned from some basic facts about multi-criteria optimization (well-known for many participants of our seminar) and introduced various multi-criteria optimization problems found in semiconductor manufacturing to the participants of the other seminar.

Friday consisted of a panel discussion (panelist Heavey, Lendermann, Matta, Pedrielli) moderated by John Fowler on the required core elements of a simulation testbed which allows for assessing smart planning and control decisions in the semiconductor industry and a wrap-up session.

**Table 2** Breakout Sessions.

| Session | Topic | Participants (lead in bold) |
|---|---|---|
| 1 | What are the state-of-the-art and future needs for ML-based decision support for different (smart) planning and control tasks | Ehm, Hautz, **Jang**, Lendermann, Rose, Schmielau, Stehli |
| | same topic | Bitsch, Dauzère-Pérès, **Heavey**, Korde, Mönch, Prisacaru, Pedrielli |
| | same topic | Bisslich, Fowler, Hassoun, Matta, **Ponsignon**, Schorn, Shekari, Shen |
| 2 | What are the state-of-the-art and future needs for simulation-based decision support for different (smart) planning and control tasks? | Bisslich, Hautz, Jang, **Matta**, Mönch, Ponsignon, Prisacaru, Stehli |
| | same topic | Bitsch, Ehm, Dauzère-Pérès, **Hassoun**, Korde, Pedrielli, Shekari |
| | same topic | Fowler, **Heavey**, Lendermann, Rose, Schmielau, Schorn, Shen |
| 3 | Sustainability issues | Bitsch, Dauzère-Pérès, Hassoun, **Mönch**, Schorn |
| | Information systems and ontologies | **Ehm**, Matta, Prisacaru, Schmielau, Shekari, Stehli |
| | Preparation of exchange with the participants of the parallel seminar on multi-objective optimization under budget | Bisslich, **Fowler**, Ponsignon |

## 3.2 Key Take Aways

There were a number of key findings and areas for future research that were identified in the seminar. We will first summarize some of the key findings and will follow this with some areas for future research.

One of the first findings was that the participants generally agreed that some of the major elements of smart manufacturing are already implemented in semiconductor manufacturing, but there are also elements that are less well understood and consequently implemented. Having said this, ML approaches are considered as promising for semiconductor manufacturing, but their potential is still not fully understood and explored. This is especially true for the role of reinforcement learning which is recently often applied to semiconductor scheduling.

Second, it appears that there are still limitations in applying different simulation paradigms in practice such as ML approaches are often not integrated into simulation models and existing reference simulation models are too difficult to apply for benchmarking purposes. Digital twins are considered as another promising direction for semiconductor manufacturing, however, they are not fully implemented so far.

Third, both the industrial and academic participants generally agree that the integration of sustainability efforts into decisions on the wafer fab and the supply chain level is often fairly ad hoc and could/should be improved in the future.

Finally, the participants generally agreed that there does not currently exist an adequate reference (simulation) model for smart manufacturing. Such a simulation model should allow for making sustainable decisions and for supporting the application of ML techniques when decisions are made.

In addition to the findings mentioned above, several areas for future research were identified.

An overarching idea was that the future research should focus more on formulation of appropriate models for smart manufacturing because this is fundamentally more important than the actual solution techniques chosen. Some of the future research areas are included below:

- More applications of CPSs, for instance for AMHS operations or for lot processing are desirable in semiconductor manufacturing.
- Multi-agent systems (MAS) are a desirable software paradigm for smart semiconductor manufacturing. However, more real-world applications are required.
- Developing better integration of various decisions made in the elements of Figure 1.
- Incorporating sustainability aspects into strategic and tactical supply chain planning models.
- Exploring the use of different simulation paradigms (systems dynamics, agent-based, hybrid models, reduced simulation models) to model and analyze semiconductor supply chains.
- Exploring and applying the possibilities of the semantic web to facilitate a meaningful data exchange between different planning and control applications.

## 3.3 Next Steps

As a way to further the discussion of and collaboration on the topics of the seminar, Prof. Lars Mönch, Hans Ehm, and Prof. John Fowler are guest editing a special issue of the *Flexible Services and Manufacturing Journal* entitled *Decision-Making Techniques for Smart Semiconductor Manufacturing*. The deadline for submission is May 1, 2024. This date was selected to allow time for ideas created by the participants of the seminar to be incorporated into papers `https://www.springer.com/journal/10696/updates/26269410`.

## 3.4 Acknowledgements

First, we want to thank the Dagstuhl staff for their great support of this seminar. The seminar also would not have been nearly as productive without the active contribution of every attendee, and for that the organizers are extremely grateful.

**References**

1. Chien, C.-F, Chuang, S.-C. (2014): A Framework for Root Cause Detection of Sub-Batch Processing System for Semiconductor Manufacturing Big Data Analytics. *IEEE Transactions on Semiconductor Manufacturing*, 27(4), 475–488.
2. Chien, C.-F., Dauzère-Pérès S., Ehm H., Fowler J. W., Jiang Z., Krishnaswamy S., Lee, T.-E., Mönch, L., Uzsoy, R. (2011): Modeling and Analysis of Semiconductor Manufacturing in a Shrinking World: Challenges and Successes. *European Journal of Industrial Engineering*, 5(3), 254–271.
3. Chien, C.-F., Ehm, H., Fowler, J. W., Mönch, L., Wu, C.-H. (2021): Production-Level Artificial Intelligence Applications in Semiconductor Manufacturing. *Proceedings of the 2021 Winter Simulation Conference*, 1–8.
4. Chien, C.-F., Ehm, H., Fowler, J. W., Kempf, K. G., Mönch, L., Wu, C.-H. (2023): Production-Level Artificial Intelligence Applications in Semiconductor Manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 36(4), 560–569.
5. Dequeant, D., Vialletelle, P., Lemaire, P., Espinouse, M.-L. (2016): A Literature Review on Variability in Semiconductor Manufacturing: The Next Forward Leap to Industry 4.0. *Proceedings of the 2016 Winter Simulation Conference*, 2598–2609.

6 Ehm, H., Mönch, L. (2016): Von der Digitalen Fabrik zur Digitalen Lieferkette in der Halbleiterindustrie: Bestandsaufnahme, Lösungsansätze und viele Herausforderungen. Obermaier, R. (ed.): *Industrie 4.0 als unternehmerische Gestaltungsaufgabe*, Springer/Gabler, 149–158.

7 Forstner, L., Dümmler, M. (2014): Integrierte Wertschöpfungsnetzwerke – Chancen und Potenziale durch Industrie 4.0. *Elektrotechnik und Informationstechnik*, 131(7), 199–201.

8 Ham, A., Kim, D. J. (2017): Scheduling of Drone-based Material Transfer System in Semiconductor Manufacturing. *Proceedings of the 2017 Winter Simulation Conference*, 3726–3732.

9 Hwang, I., Jang, Y. J. (2020): Q($\lambda$) Learning-based Dynamic Route Guidance Algorithm for Overhead Hoist Transport Systems in Semiconductor Fabs. *International Journal of Production Research*, 58(4), 1199–1221.

10 Kaufmann, T., Forstner, L. (2017): Die horizontale Integration der Wertschöpfungskette in der Halbleiterindustrie – Chancen und Herausforderungen. T. Bauernhansl, M. ten Hompel, B. Vogel-Heuser (eds.): *Industrie 4.0 in Produktion, Automatisierung und Logistik*, Wiesbaden, 359–367.

11 Khakifirooz, M., Chien, C.-F., Chen, Y.-J. (2018): Bayesian Inference for Mining Semiconductor Manufacturing Big Data for Yield Enhancement and Smart Production to Empower Industry 4.0. *Applied Soft Computing*, 68, 990–999.

12 Khakifirooz, M., Fathi, M., Wu, K. (2019a): Development of Smart Semiconductor Manufacturing: Operations Research and Data Science Perspectives. *IEEE Access*, 7.

13 Khakifirooz, M., Fathi, M., Chien, C.-F., Pardalos, P. M. (2019b): Management Suggestions for Process Control of Semiconductor Manufacturing: An Operations Research and Data Science Perspective. Blondin, M. J., Pardalos, P. M., Sáez, J. S. (eds.) *Computational Intelligence and Optimization Methods for Control Engineering*, Springer, 245–274.

14 Kopp, D., Hassoun, M., Kalir, A., Mönch, L. (2020): SMT2020 – A Semiconductor Manufacturing Testbed. *IEEE Transactions on Semiconductor Manufacturing*, 33(4), 522–531.

15 Kusiak, A. (2018): Smart Manufacturing. *International Journal of Production Research*, 56(1-2), 508–517.

16 Kusiak, A. (2019): Fundamentals of Smart Manufacturing: A Multi-thread Perspective. *Annual Reviews in Control*, 47, 214–220.

17 Lee, J. (2015): Smart Factory Systems. *Informatik-Spektrum*, 38(3), 230–235.

18 Mertens, P. (2014): Industrie 4.0 = CIM 2.0? *Industrie Management*, 30(1), 27–30.

19 Mertens, P., Barbian, D., Baier, S. (2017): Digitalisierung und Industrie 4.0 – eine Relativierung. Springer Vieweg, Wiesbaden.

20 Mertens, P., Barbian, D. (2019): Industrie 4.0 und Integrierte Informationsverarbeitung. Obermaier, R. (ed.): *Handbuch Industrie 4.0 und Digitale Transformation*, Springer/Gabler, 47–71.

21 Mönch, L., Fowler, J. W., Mason, S. J. (2013): *Production Planning and Control for Semiconductor Wafer Fabrication Facilities: Modeling, Analysis, and Systems*. Springer Operations Research/Computer Science Interfaces, New York, Vol. 52.

22 Rocholl, J., Mönch, L., Fowler, J. W. (2020): Bi-criteria Parallel Batch Machine Scheduling to Minimize Total Weighted Tardiness and Electricity Cost. *Journal of Business Economics*, 90, 1345–1381.

23 Rocholl, J., Mönch, L. (2021): Simulation-Based Performance Assessment of Sustainable Manufacturing Decisions. *Proceedings of the 2021 Winter Simulation Conference*, 1–12.

24 Sucky, E., Gampl, M., Ruh, A., Stelzer, N., Weidinger, J. (2018): Industrie 4.0: Marketing-kampagne oder Revolutionsbeginn? `https://opus4.kobv.de/opus4-bamberg/-frontdoor/index/index/docId/46573`.

# 4 Overview of Talks

## 4.1 Complex Flexible Job-shop Scheduling Problems and Semiconductor Manufacturing

*Stéphane Dauzère-Pérès (Mines Saint-Etienne, FR)*



The presentation focused on the flexible job-shop scheduling problem and some of its extensions, on neighborhood-based metaheuristics and their application to semiconductor manufacturing. The extensions include additional constraints, such as sequence-dependent setup times and batching, and new criteria. The disjunctive graph modeling was presented for various complex flexible job-shop scheduling problems, together with various properties that allow to speed up the search in neighborhhod-based metaheuristics. Scheduling in semiconductor manufacturing was then discussed to emphasize the complexity and the size of the real-life problems that must be solved. A batch-oblivous approach was then presented, which has been implemented and is being used in a real factory to solve problems with more than 2,500 operations and 200 machines. The presentation ended with some general conclusions, in particular on various industrial constraints not discussed but already considered, and perspectives on future relevant academic and industrial research, in particular robust scheduling.

## 4.2 Reducing Bullwhip in Supply Chains Containing Semiconductors Using Anonymous Survey and Semantic Web Technologies

*Hans Ehm (Infineon Technologies – München, DE)*


**URL** https://www.researchgate.net/lab/Supply-Chain-Innovation-Hans-Ehm

Corona triggered a bullwhip amplified demand reduction for semiconductors – especially from the automotive industry. In conjunction with rising demand in communication industry (more home offices, more cameras, more audio tools, ...) global capacities for the automotive industry was lost and this caused global chip shortages when the demand came back in the automotive industry with the consequence of shutting off of car manufacturing factories.

With analytics and simulation the root causes of the root causes could be identified, which is beyond others a Kanban driven replenishment which acts a bullwhip accelerator in disruptive times like during COVID and the human behavior based on prospect theory.

Due to the magnitude and possible further impacts of the problem not only the usual business stakeholders searched for long term solution also governments got involved and triggered and supported decision techniques to learn from the problem and mitigate it for the future.

The 4-step plan emerged: 1) higher inventory, 2) anonymous survey and 3) breakdown this coarse survey results with 3) semantic web based AI techniques to enable a 4) a leadtime based pricing. This plan is now in implementation in EU funded projects and supported by semi IAC

Furthermore general current problems and opportunities in the domain semiconductor and supply chains containing semiconductors have been shared from an industry point of view.

In the URL noted above full text of around 100 papers relevant for the domain (not limited to the presentation title), are provided.

## 4.3 Scheduling in Semiconductor Reliability Testing Labs

*Jessica Hautz (KAI – Villach, AT) and Lars Mönch (FernUniversität in Hagen, DE)*



Power semiconductor components have to fulfill high-quality standards. To meet these requirements, so-called Reliability Product Testing (RPT) laboratories perform product qualification tests, process monitoring tests, and tests accompanying the technology development. Reliability testing is resource-intensive, requiring trained engineers and high-tech equipment. The complex allocation of tests to the respective resources to create a scheduling plan within RPT labs is a very challenging task. Currently, this plan is created by senior experts with the help of a static dispatcher that doesn't consider resource capacities. Introducing scheduling models to the RPT labs has the potential to reduce equipment idle times, to avoid bottlenecks, to meet the costomer deadlines more confident, and to use resources more efficiently. Also, costly lab extensions and equipment purchases can be prevented. The problem we are considering belongs to the family of complex job shop scheduling problems, having the following $\alpha|\beta|\gamma$-representation:

$$FJ_m \mid aux, prec, reentr, w_j, r_j, d_{ji}, t_{jik}, p-batch, B_k, s_j, s_f, pmtn, s-batch, s_{ji\ell k}, incompatible \mid \sum w_j T_{ji} + \sum w_j T_{jik} + \alpha \cdot C_{max}.$$

The development of a rolling-horizon approach to solve the problem using metaheuristics based on disjunctive graphs will be investigated.

## 4.4 Demonstration of the Feasibility of the Application of Machine Learning for Production Scheduling

*Cathal Heavey (University of Limerick, IE)*



**Joint work of** Amir Ghasemi, Amir Ashoori, Cathal Heavey
**Main reference** Amir Ghasemi, Amir Ashoori, Cathal Heavey: "Evolutionary Learning Based Simulation Optimization for Stochastic Job Shop Scheduling Problems", Appl. Soft Comput., Vol. 106, p. 107309, 2021.
**URL** https://doi.org//10.1016/J.ASOC.2021.107309

Simulation Optimization (SO) techniques refer to a set of methods that have been applied to stochastic optimization problems, structured so that the optimizer(s) are integrated with simulation experiments. Although SO techniques provide promising solutions for large and complex stochastic problems, the simulation model execution is potentially expensive in terms of computation time. Thus, the overall purpose of this research is to advance the evolutionary SO methods literature by researching the use of metamodeling within these techniques. Accordingly, we present a new Evolutionary Learning Based Simulation

Optimization (ELBSO) method embedded within Ordinal Optimization. In ELBSO a Machine Learning (ML) based simulation metamodel is created using Genetic Programming (GP) to replace simulation experiments aimed at reducing computation. ELBSO is evaluated on a Stochastic Job Shop Scheduling Problem (SJSSP), which is a well known complex production planning problem in most industries such as semiconductor manufacturing. To build the metamodel from SJSSP instances that replace simulation replications, we employ a novel training vector to train GP. This then is integrated into an evolutionary two-phased Ordinal Optimization approach to optimize an SJSSP which forms the ELBSO method. Using a variety of experimental SJSSP instances, ELBSO is compared with evolutionary optimization methods from the literature and typical dispatching rules. Our findings include the superiority of ELBSO over all other algorithms in terms of the quality of solutions and computation time. Furthermore, we present how approaches similar to the ELBSO method could be integrated with a Manufacturing Execution System (MES) in semiconductor manufacturing to allow scheduling at an operational level.

## 4.5 Agent-based Decision Support in Borderless Fab Scenarios in Semiconductor Manufacturing

*Raphael Herding (Westfälische Hochschule – Bocholt, DE) and Lars Mönch (FernUniversität in Hagen, DE)*



The design and the implementation of a multi-agent system (MAS) for a borderless fab scenario is presented. In a borderless fab scenario, lots are transferred from one semiconductor wafer fabrication facility (wafer fab) to another nearby wafer fab to process certain process steps of the transferred lots. Production planning is carried out individually for each of the wafer fabs. The modeling of the available and requested capacity in the production planning models of the participating wafer fabs is affected by the lot transfer. Three scenarios, namely no borderless fab (NBF), borderless fab with no production planning (BF-NPP), and borderless fab with advanced production planning (BF-APP) are discussed and the performance results are presented. The transfer of the route information from one wafer fab to another one to automatically generate the linear programming models for production planning is described. Production planning is carried out in a rolling horizon setting. We show by simulation experiments that a correct modeling of the capacity in the production planning formulations results in improved profit compared to a setting where the lot transfer is not taken into account in the planning formulations. In addition, we demonstrate that an ontology to standardize the data exchange between the wafer fabs can be beneficial in a borderless fab setting.

## 4.6 Advancing Automation Through Robot Collaboration Intelligence and Digital Twin Integration

*Young Jae Jang (KAIST – Daejeon, KR)*



This presentation delves into the evolving realm of automation, emphasizing the role of robot collaboration intelligence. Leveraging advanced AI, this innovation facilitates synergistic interactions among industrial robots, streamlining their control and management. With the rise of adaptable agent-based robots, such as automated guided vehicles (AGVs), autonomous mobile robots (AMRs) and overhead hoist transporters (OHTs) in manufacturing, there's a marked increase in operational flexibility. As these robots grow more sophisticated and their numbers expand, robot collaboration intelligence emerges as a pivotal tool, amplifying their efficacy. Through industry case studies, we will elucidate the immense potential of this nascent technology:

- Management of over 1,000 Overhead Hoist Transport (OHT) Systems in semiconductor fabrication plants
- Fleet regulation of 200 AGV/AMR in warehouse settings

Additionally, we introduce the Digital Twin concept tailored for Robot Collaboration Intelligence. The outlined Digital Twin (OMS-DT) encompasses a robot emulator, integrated hardware-software links, and a simulated environment, with real-world applications of this Digital Twin being showcased.

## 4.7 On the Application of Machine Learning in Semiconductor Manufacturing

*Adar Kalir (Intel Israel – Qiriat-Gat, IL)*



Semiconductor manufacturing is data-intensive. It is also very complex. Combined, these two aspects drive ML usage in this industry. Big success is already evident in Yield, Equipment Diagnosis, ... and growing in productivity, capacity. "Signal-to-Noise" is still a challenge in many problems (e.g. wafer breaks; CQT's [violations]).

## 4.8 Minimizing Makespan for a Multiple Orders per Job Scheduling Problem in a Two-stage Permutation Flowshop

*Rohan Korde (Arizona State University – Tempe, US), John Fowler (Arizona State University – Tempe, US), and Lars Mönch (FernUniversität in Hagen, DE)*



We introduced the "multiple orders per job" scheduling problem (Mason et al., 2004) in a two-stage permutation flowshop with the goal of minimizing the makespan. This problem, $F2\|\ moj(.), Bk(.), prmu\|\ C_{max}$ is NP-hard. We discussed different types of methods we used to solve this problem for both small-sized and large-sized problem instances. These methods include exact methods, heuristics, and metaheuristics. Finally, we compared the performance of these methods on the different problem instances that were generated using a full factorial numerical experiment.

## References


1 Graham, R. L., E. L. Lawler, J. K. Lenstra, and A. H. G. Rinnooy Kan. 1979. "Optimization and Approximation in Deterministic Sequencing and Scheduling: A Survey". Annals of Discrete Mathematics 5:287–326.

2 Johnson, S. M. 1954. "Optimal Two and Three-Stage Production Schedules with Setup Times Included", Naval Research Logistics Quarterly 1:61-67.

3 Laub, J. D., J. W. Fowler, and A. B. Keha. 2006. "Capacitated Lot Streaming in 3-Machine Flow Shops", ASU Working Paper ASUIE-ORPS-2006-016.

4 Laub, J. D., J. W. Fowler, and A. B. Keha. 2007. "Minimizing Makespan With Multiple-Orders-Per-Job in a Two-Machine Flowshop", European Journal of Operational Research 182(1):63-79.

5 Mason, S. J., P. Qu, E. Kutanoglu, and J. W. Fowler. 2004. "The Single Machine Multiple Orders Per Job Scheduling Problem", Working paper.

6 Nawaz, M., E. E. Enscore, Jr, and I. Ham. 1983. "A Heuristic Algorithm for the m-Machine, n-Job Flowshop Sequencing Problem", Omega – the International Journal of Management Science 11:91-95.

7 Palmer, D. S. 1965. "Sequencing Jobs Through a Multi-Stage Process in the Minimum Total Time – A Quick Method of Obtaining a Near Optimum", Operational Research Quarterly 16:101-107.

8 Pinedo M. L. 2022. Scheduling Theory, Algorithms, and Systems. 6th ed. Cham, Switzerland: Springer Nature Switzerland AG.

9 Sarin, S. C., L. Wang, and M. Cheng. 2012. "Minimising makespan for a two-machine, flow shop, single-wafer-processing, multiple-jobs-per-carrier scheduling problem", International Journal of Planning and Scheduling 1(3):171-208.


## 4.9 Smart Capacity Planning and Material Flow Optimisation in the Semiconductor Wafer Fab

*Peter Lendermann (D-SIMLAB – Singapore, SG)*



In the "AI-Driven Smart Factory" as propagated by SEMI in the Smart Manufacturing Roadmap, Artificial Intelligence, and in particular Reinforce Learning techniques are a distinguishing enabling element along the transition from today's Smart/Industry 4.0 approaches to the future "Smart 2.0 Factory" and especially the "Autonomous Control Room". This raises the question for what kind of tasks and functions such techniques can add value on top of what is possible with established methods, which in this presentation will be discussed from the Industrial Engineering point of view, i.e. decisions around capacity planning and material flow optimisation in a semiconductor wafer fab. Whenever such a decision is to be taken, the expected performance of the selected option with regard to a particular objective against alternative options is required. In the case of a wafer fab, the underlying objective function should be a capacity model that is able to represent a Discrete Event Logistics System, and since in most cases the interdependency between capacity and cycle time, i.e. the causality between a certain solution option and its impact on capacity and cycle time needs to be considered, this capacity model should be (in the case of a schedule to be generated) a deterministic or (in the case of a longer term plan to be generated) a stochastic Discrete Event Simulation model, also because these causalities need to be portrayed with sufficient fidelity.

To determine an optimal plan or schedule, powerful optimisation techniques are needed, requiring detection of correlations between decision variable values and objective values. That is where Reinforcement Learning techniques come into the picture. The presentation will also explain why such Reinforcement Learning techniques can be more useful for scheduling but may have limited value for situation-based dispatching. Also, in an environment where the underlying capacity model is always a simplified representation of actual operations, an optimal solution would never be found anyway. Rather, for all practical purposes it is sufficient to determine a much better solution with as few iterations as possible.

## 4.10 Smart (Semiconductor) Manufacturing

*Lars Mönch (FernUniversität in Hagen, DE) and John Fowler (Arizona State University – Tempe, US)*



In this talk, we discuss the notion of smart (semiconductor) manufacturing. The talk is motivated by the fact that the notion of smart manufacturing is fuzzy to some extent. The term is often used by software vendors. Goal of this presentation: some unification as a prerequisite for the seminar based on the diverse existing literature. The roots of Smart Manufacturing can be find in flexible manufacturing systems, computer integrated manufacturing (CIM), and the intelligent manufacturing systems (IMS) program. We start by defining the term smart manufacturing. The pillars materials, manufacturing technology processes, data, predictive manufacturing, and sustainability will be discussed. Opportunities and challenges for smart manufacturing are described. Moreover, applications to semiconductor manufacturing are described. Research topics for semiconductor smart manufacturing are identified in the last part of the talk.

## 4.11 DTFab: Performance Improvement, Analytics and Security in DT – controlled Semiconductor Systems

*Giulia Pedrielli (Arizona State University – Tempe, US)*



Digital factories have been recognized as a paradigm with considerable promise for improving manufacturing performance. Digital Twins have emerged as a powerful tool to improve control performance for large-scale smart manufacturing systems. We argue that DT-based smart factories are vulnerable to attacks that use the DT to damage the system while remaining undetectable, specifically in high-cost processes, where DT technologies are more likely to be deployed. As an instructive example, we look into smart semiconductor processes with focus on photolithography.

We formulate a static optimization problem to maximize the damage of a cyber-attack against a photolithography digital twin that minimizes detectability to the process controller. Results demonstrate that this problem formulation provides attack policies that successfully reduce the throughput of the system at trade off of increased detectability to a common process control technique. Results encourage more research in the domain, especially to face scalability and policy-like solutions.

## 4.12 Consideration of Customer Agreements in End-to-End Supply Chain Planning Processes

*Thomas Ponsignon (Infineon Technologies – München, DE)*



Customer agreements have gained increasing strategic importance in the semiconductor business over the last two years. One type of customer agreements is Capacity Reservation Agreement (CRA). CRA is a premium service offered to selected customers with key benefits for both the manufacturer and the customer. However, supply chains are required to handle not only CRAs, but a large variety of customer agreements, which bring along new problem statements during the entire contract lifecycle. There are only very few papers in the literature addressing those challenges. Infineon's approach was to implement a dedicated application called Customer Agreement Tracking Solution (CATS), which allows the centralized and digitalized storage of volume-related customer agreements as a single source-of-truth system. CATS data is used throughout all phases, from contract negotiation to contract fulfilment, to support decision-making in downstream supply chain processes and systems. Further details about Infineon's practices are provided along long-, mid-, and short-term decisions. Remaining open challenges are outlined. Finally, implications for smart semiconductor manufacturing are described along the lines of horizontal and vertical integration.

## 4.13 Semiconductor Manufacturing Digitalisation Challenges

*Alexandru Prisacaru (Bosch GmbH – Stuttgart, DE)*



In semiconductor manufacturing, many process steps are required to realize the desired chip design. Up to 600 steps and five months are needed to produce a wafer. Complexity further increases when having a high volume and high mix of products. Digitalisation is helping in handling such complexity to improve production and optimize utilization. Digitalisation comes with different challenges that must be tackled. Data availability and connectivity of the production systems are only sometimes available for the analytics. The data is either distributed in silos or in closed systems. In addition, the existing databases are designed for local purposes and do not have big data functionalities in the design. Another challenge represents the equipment legacy, in which software or hardware updates can be costly to have the required functionality. Software extensions built internally must have additional infrastructure to support it. In a manufacturing environment, software reliability plays an important role. It must always be available and robust enough to fulfill the high requirements. Most proof of concept fails because it does not bring any advantage or is costly. Additionally, finding the balance between a one-solution/one-platform strategy and a very complex and hard-to-train many-solution platform strategy is challenging. Even when these challenges are met and solved, there is the human factor challenge, mainly how to digitize the expert knowledge. Highly trained teams with different skill sets are required to operate these solutions.

## 4.14 Cloud-based Simulation Experiments for Optimization and Machine Learning

*Oliver Rose (Universität der Bundeswehr – München, DE)*


**Main reference** Warren B. Powell: "Reinforcement Learning and Stochastic Optimization: A Unified Framework for Sequential Decisions", John Wiley & Sons, Ltd, 2022
**URL** https://doi.org/10.1002/9781119815068

In simulation-based decision support in industry and the military, large numbers of configurations are usually examined using simulation experiments. This is true for classical optimization problems of operations research as well as for the creation of machine-learned models, especially for decision making based on reinforcement learning. In both cases, enormous amounts of simulation data are generated and consumed. Running the experiments on ordinary office computers is possible in principle but leads to exceptionally long problem-solving times. It therefore makes sense to outsource the experiments to a "simulation cloud". The paper/chapter deals with the software and hardware requirements for such a simulation infrastructure and gives first insights into the expected performance of this solution.

## 4.15 Genetic Programming for Energy-aware Scheduling

*Daniel Schorn (FernUniversität in Hagen, DE) and Lars Mönch (FernUniversität in Hagen, DE)*



We consider a scheduling problem for identical parallel batch processing machines in semiconductor wafer fabrication facilities (wafer fabs). Unequal ready times of the jobs and incompatible job families are assumed. An integrated objective function consisting of the total weighted tardiness and the total electricity cost is considered. A time-of-use (TOU) tariff is assumed. A genetic programming procedure is proposed to automatically discover dispatching rules for list scheduling approaches. A decision theory heuristic is used to decide when to schedule idle times on the machines to improve the TEC measure. A time window decomposition is applied to take into account the different ready times of the jobs. Results of the computational experiments show that the learned dispatching rules lead to high-quality schedules in respect to the integrated objective function.

## 4.16 Order-lot Pegging in a Multi-Fab Setting

*Liji Shen (WHU – Vallendar, DE), John Fowler (Arizona State University – Tempe, US), Lars Mönch (FernUniversität in Hagen, DE)*



This study considers the highly sophisticated wafer fabrication process and extracts a specific lot-order pegging problem. Given are a set of orders containing different numbers and types of wafers, as well as multiple identical wafer fabs. Wafer lots in each fab which are either already released or to be released, are used to satisfy orders. Our goal is to find a matching of orders and lots so that the total tardiness is minimized.

In our approach, we first consider genetic algorithm (GA) which is based on the random-key representation for parallel machine problems. Besides the population-based method, we develop an iterated local search (ILS) algorithm with diverse rules for generating initial solutions and determining fab assignment, as well as neighbourhoods. To further improve performance, we also combine GA and ILS.

For our computational tests, we adopt and extend the problem instances in the literature where a simulated annealing (SA) algorithm is proposed. When solving the original instances with one fab, ILS outperforms SA. ILS also reaches better solutions comparing to GA in a multi-fab setting. On the other hand, hybridizing GA and ILS achieves best results.

## 4.17 Automation Challenges in Semiconductor Fabs

*Marcel Stehli (Globalfoundries – Dresden, DE)*



Semiconductor manufacturing is one of the most complex production systems ever build. To manage the complexity, enormous efforts have been made in the last decades to automate the manufacturing process. Nowadays for modern 300mm manufacturing lines the entire value creation chain within a fab is fully automated. This concerns the actual wafer processing as well as the logistics around the production process. Standardization of machine interfaces, transport systems and software interface have made a decisive contribution here.

The achivements in full fab automation provide a significant baseline for the overall optimization of the production process from a WIP flow perspective. However, challenges remain due to the systems complexity and interdependcies and the still required humand interaction with the system and its processes. Furthermore it remains challenging to build an overall and comprehensive fab control model that allows for a unified and standardized way of controlling and optimization of the fab operation.

## 4.18 Machine Learning-based Process Modeling in Semiconductor Manufacturing

*Gian Antonio Susto (University of Padova, IT)*


**Main reference** Mattia Carletti, Matteo Terzi, Gian Antonio Susto: "Interpretable Anomaly Detection with DIFFI: Depth-based feature importance of Isolation Forest", Eng. Appl. Artif. Intell., Vol. 119, p. 105730, 2023.
**URL** https://doi.org//10.1016/J.ENGAPPAI.2022.105730

Machine Learning (ML) has had a tremendous impact on many industries, especially those that rely heavily on data. In Semiconductor Manufacturing (SM), ML solutions have proven highly effective at various levels. The seminar primarily focused on ML applications at the process level, involving sensors and metrology data. Several technologies fall within this domain, including Predictive Maintenance, Virtual Metrology, Fault Detection, Dynamic Sampling, to name just a few.

The seminar emphasized Anomaly Detection, a critical task in process-level modeling aimed at enhancing monitoring capabilities. Currently, the industry relies mainly on univariate control charts for process monitoring, which have several limitations: (i) they fail to capture the multivariate nature of the problem; (ii) they rely on unimodal/Gaussian distribution assumptions; (iii) they can overwhelm users when monitoring hundreds of Key Performance Indicators (KPIs). Unsupervised Anomaly Detection (AD) tools can address these issues, offering comprehensive and concise information that can be utilized in Decision Support Systems.

During the seminar, we also introduced DIFFI, an approach designed to make Isolation Forest, arguably the most popular and effective AD method, more interpretable. This has the potential to significantly impact the adoption and trustworthiness of AD, as well as expedite root cause analysis and decision-making processes.

Furthermore, we highlighted some general challenges in developing ML solutions for semiconductor manufacturing and identified potential actors who could successfully overcome these obstacles.

# 5 Breakout Reports

## 5.1 Breakout Session 1a

*Young Jae Jang (KAIST – Daejeon, KR), Henrik Schmielau (Infineon Technologies – München, DE), Hans Ehm (Infineon Technologies – München, DE), Jessica Hautz (KAI – Villach, AT), Peter Lendermann (D-SIMLAB – Singapore, SG), Oliver Rose (Universität der Bundeswehr – München, DE), and Marcel Stehli (Globalfoundries – Dresden, DE)*



### Quality Control

Supervised learning has become a cardinal tool in overseeing defect identification processes, fundamentally altering the traditional approaches witnessed in semiconductor manufacturing environments. Previously, the identification and sorting of defected chips or wafers hinged on manual visual inspection or rudimentary rule-based classification systems. Leveraging the data accrued from these past exercises, contemporary machine learning algorithms are trained to facilitate an automated, and rapid identification and sorting of defects. A notable method employed in this endeavor is deep learning predicated on supervised learning paradigms.

### Machine Maintenance

Supervised learning is once again pivotal in discerning abnormalities in machine function. Notwithstanding the prevalence of supervised learning, it is faced with impediments such as the onerous task of data labeling delineating normal and abnormal functionalities, compounded by frequent unavailability of requisite data. This has engendered a transition towards unsupervised learning approaches, where strategies like deep auto-encoders are gaining traction owing to their aptitude in addressing the aforementioned challenges effectively.

### Operations Decision-Making

A critical facet of semiconductor fabrication operations encompasses a broad spectrum of activities including, but not limited to, inventory control, lot release control, and production planning and scheduling. Equally imperative are the strategies deployed in lot dispatching and

Overhead Hoist Transfer (OHT) vehicle routing and dispatching. In recent times, there has been a discernible shift towards adopting reinforcement learning approaches, diverging from the traditional optimization or rule-based decisions that previously held sway. For instance, many large chip manufacturers have embraced reinforcement learning in OHT dispatching and routing. Parallelly, processing machine scheduling and dispatching are gradually being recalibrated to incorporate data and simulation-based optimization leveraging the strengths of reinforcement learning. This trend signals a commitment to efficiency and optimization, utilizing the capabilities of modern learning approaches to enhance decision-making processes.

This in-depth analysis showcases the modern advancements in the manufacturing sector, accentuating the pivotal role of learning algorithms in steering quality control, machine maintenance, and operational decision-making towards an epoch of heightened efficiency and precision. It is pertinent to continue exploring and expanding upon these technological strides to foster a robust manufacturing landscape that is aligned with the demands of contemporary production exigencies.

### Future needs

In consideration of imminent advancements in artificial intelligence (AI) applications within the manufacturing domain, a pivotal focus rests on fostering a synergistic collaboration between AI specialists and domain experts. The deployment of AI cannot reach its optimum efficacy without the integral guidance of experienced field engineers in data labeling processes, ensuring a judicious application of AI mechanisms rooted in a profound understanding of both the algorithmic logic and the relevant domain knowledge.

Given the inherent complexity of AI algorithms, which often preclude a comprehensive understanding of their underlying logic, it becomes incumbent upon decision-makers to at least grasp the foundational principles of AI model construction. This facilitates more informed decisions, precluding misconceptions that may arise from a lack of understanding of the modeling assumptions and the provenance of the data utilized. Consequently, we underscore the necessity of collaboration between AI and domain experts to mitigate potential errors in conclusions derived from AI analytics.

In furtherance of this objective, we propose the integration of expert systems that amalgamate conventional knowledge and logical reasoning with machine learning. This approach promises to be a robust conduit for the cohesive assimilation of domain and AI expertise, generating solutions that are both innovative and grounded in established knowledge.

Moreover, we envision the establishment of a digital test bed, operationalizing a virtual factory predicated on simulated factory operations, to serve as a fertile ground for AI and domain experts to foster collaborative innovations. This virtual environment not only facilitates a harmonious collaborative endeavor but also engenders a repository of simulated data, enhancing the reliability and efficacy of AI applications in real-time operations.

Lastly, we turn our focus to the burgeoning application of reinforcement learning in Automated Material Handling System (AMHS) operations. Despite its nascent stage in acceptance, it harbors considerable potential to revolutionize other operational realms including inventory control, lot release control, and lot dispatching. We posit that these domains present a ripe landscape for the efficacious implementation of machine learning solutions predicated on reinforcement learning, thereby heralding a new frontier in manufacturing optimization.

In conclusion, as we stand on the cusp of a transformative era in manufacturing, it is imperative to strategically leverage the collaborative might of AI and domain expertise, steering towards a future characterized by innovation, efficiency, and precision. The propositions delineated herein advocate for a meticulous and collaborative approach to integrating AI in manufacturing, urging an embracing of systems that are both logical and adaptive, with an eye towards a harmonized and prosperous future in manufacturing.

## 5.2 Breakout Session 1b

*Cathal Heavey (University of Limerick, IE), William Bitsch (WHU – Vallendar, DE), Stéphane Dauzère-Pérès (Mines Saint-Etienne, FR), Lars Mönch (FernUniversität in Hagen, DE), Giulia Pedrielli (Arizona State University – Tempe, US), and Alexandru Prisacaru (Bosch GmbH – Stuttgart, DE)*



After initial discussion a number of questions were derived among the breakout group. These questions are given below:

What are the **problems that ML can tackle in semiconductor topics:**

- Root cause analysis: take measures for specific lots and see the impact on the yield of the different parts
- Many applications are now in the control part, examples are:
  - Sampling of the lot: how do one choose if a lot is important to measure a lot or not
  - Process control
    - Virtual metrology:
      - Predict the value of the measure. Try to predict the value of the metric
      - measurements
      - SPC enhancement, with techniques that replace the traditional SPC with control charts
      - in general advanced process control (APC)
      - equipment Process Control: controlling the machining parameters
      - run-to-run: take measures and change machining parameters and ML is used to build the run-to-run model
  - Production control

  Cycle time prediction is another important area: ML is paired with the MES system in this case. It could be:
  - Fab level:
  - Single step:
  - Phases in the fab that predict where the lot is going to be in the next few days: this can be for:
    - Production planning
    - Scheduling
- Dispatching & Scheduling (ML is used more towards dispatching)
- Lot release/order release could also be done via ML
- ML to shorten simulation time, through the use of a metamodel
- Optimize parameters of a dispatching rule to calibrate the hyperparameters of the policy
- Predictive Maintenance:
  - ML can be used to predict the failure time
  - Policy improvement can be used in the context of Reinforcement Learning

- Optimizing transportation within the material handling systems:
  - Evaluation of vehicle health
  - Routing of vehicles in the system

**Critical aspects**

- How do we understand what data need to be "forgotten" and what new data should be included?
- When decisions are complicated the effectiveness of the methods decreases
- Dispatching seems doable but more challenges appear for things like scheduling
- High dimensionality is still a challenge. We need to understand what the dimensionality is where these techniques work.

We also discussed questions on ML approaches, such as:

**ML methods**

- Image recoginition
- Image processing techniques
- Dynamic Neutral Network
- Large Language Models (?)
- Bayesian NN

**Challenges**

- Computing infrastructure required, what is the economical cost, what are solutions for HW architecture? Is that a huge beyond the computing power of Fab companies, as they do not want to use cloud computing
- Preparation effort:
  - Labeling
  - Annotation
  - Prepping the problem in a way that it can be handled with ML techniques
- Transparency-inexplicability
- Verification and guarantees associated to the model
- Reboustness to:
  - Changes in the features
  - Dynamical changes in time

**Other Issues**

- Explainable ML/AI – have an alternative model
- Move of expertise to software from Industrial Engineering (IE)
- Job security.

## 5.3 Breakout Session 1c

*Thomas Ponsignon (Infineon Technologies – München, DE), Dominik Bisslich (Infineon Technologies AG – Neubiberg, DE), John Fowler (Arizona State University – Tempe, US), Michael Hassoun (Ariel University, IL), Andrea Matta (Polytechnic University of Milan, IT), Daniel Schorn (FernUniversität in Hagen, DE), Mahsa Shekari (Polytechnic University of Milan, IT), and Liji Shen (WHU – Vallendar, DE)*



### Introduction

The application ML in various sectors, including smart manufacturing in the semiconductor industry, has revolutionized the way decisions are made and operations are conducted. However, this integration comes with its own set of challenges. The breakout session centered around discussions on challenges such as expertise and skill gaps, data and knowledge misalignment, standardization issues, unclear roles and expectations from ML, job security concerns, decision-making and validation challenges, and potential future directions. This essay aims to summarize the key discussions and potential solutions addressed during the session.

### Expertise and Skill Gaps

The dichotomy between AI enthusiasts and engineering domain experts is particularly evident in the semiconductor industry. This creates a perceived preference for modern ML methods over classical statistical or operational research techniques. The necessity for domain knowledge in labeling and complex decision-making was discussed, highlighting the importance of integrating domain expertise with ML knowledge to effectively address this gap in smart manufacturing.

### Data and Knowledge Gap

The misalignment between collected data and actual knowledge poses a significant challenge, impacting the transparency, verification, and robustness of ML systems in the semiconductor industry. Communication challenges between Cyber-Physical Systems (CPS) AI experts and field engineers make it difficult to understand which data can be used to solve specific problems, a crucial aspect in the complex world of semiconductor manufacturing.

The lack of established standards for ML in the manufacturing domain, including the semiconductor industry, was noted. Variances in definitions and methodologies across different vendors and suppliers, and the development of proprietary standards, question the need for universal ones. This necessitates exploring semi-standard approaches and recognizing the value-driven work of companies and the role of standards in it.

### Role and Expectations from ML

Ambiguities in setting clear goals for ML and the need for new roles like Labeler, Decision Maker, and Exception Handler were discussed in the context of smart manufacturing. There are concerns about the trustworthiness of ML and its implications, highlighting the importance of clearly outlining expectations from ML in decision support and changing classical tasks towards ML-embedded functions.

### Job Security Concerns

Fears of job displacement due to ML, along with the emphasis on the creation of advanced qualification roles, were discussed. It is crucial to recognize the shift of talents to software domains and address the talent gap in hardware sectors, including the semiconductor industry. Encouraging talents to understand both classical and ML approaches for a holistic skill set, with Factory Physics as a fundamental skill before diving into ML, is essential.

### Approach and Methodology Challenges

Balancing between supervised and unsupervised learning, the importance of labeling, and the challenges associated with it were discussed. There was a debate over the usage of images in ML modeling and their significance in semiconductor manufacturing.

### Decision Making and Validation

The challenge of verifying ML-based decisions in complex systems, like semiconductor manufacturing, and concerns about the speed versus explainability trade-off in ML were discussed. The need for human oversight in decision validation and advocacy for explainable AI to facilitate decision-making were noted. A potential solution discussed was a dual-system approach, one system for decision generation and another for explanation, which could be particularly beneficial in the complex and critical processes involved in semiconductor manufacturing.

### Applications for ML and Boundaries

Discussions included the problems that ML can tackle in the semiconductor industry, such as (Advanced) Process control, Equipment Process control, Cycle Time (CT) Predictions, Dispatching & Scheduling, Hyperparameter Tuning, Lot and order release, Predictive Maintenance, and AMHS – Routing Problems.

### Conclusions

The integration of ML into smart manufacturing in the semiconductor industry brings about various challenges. Bridging the expertise and skill gaps, addressing data and knowledge misalignment, resolving standardization issues, clarifying roles and expectations from ML, addressing job security concerns, tackling approach and methodology challenges, and ensuring proper decision-making and validation are crucial areas of focus. The potential for a dual-system approach for decision generation and explanation was discussed as a possible solution. Addressing these challenges requires a holistic approach that considers the complexities of integrating ML into the semiconductor manufacturing domain while ensuring the transparency, robustness, and effectiveness of the systems.

## 5.4 Breakout Session 2a/c

*Andrea Matta (Polytechnic University of Milan, IT), Cathal Heavey (University of Limerick, IE), Dominik Bisslich (Infineon Technologies AG – Neubiberg, DE), John Fowler (Arizona State University – Tempe, US), Jessica Hautz (KAI – Villach, AT), Young Jae Jang (KAIST – Daejeon, KR), Peter Lendermann (D-SIMLAB – Singapore, SG), Lars Mönch (FernUniversität in Hagen, DE), Thomas Ponsignon (Infineon Technologies – München, DE), Alexandru Prisacaru (Bosch GmbH – Stuttgart, DE), Oliver Rose (Universität der Bundeswehr – München, DE), Henrik Schmielau (Infineon Technologies – München, DE), Daniel Schorn (FernUniversität in Hagen, DE), Liji Shen (WHU – Vallendar, DE), and Marcel Stehli (Globalfoundries – Dresden, DE)*



Simulation-based decision support remains an integral tool for semiconductor manufacturing optimization. Using simulation-based methods can enhance operational efficiency, from individual tools to entire supply chains, fostering adaptability and strategic decision-making. Addressing current challenges in the field will help semiconductor manufacturers to improve their operations and align them with business goals, in this fast progressing industry.

### Applications in Semiconductor Manufacturing

- Semiconductor manufacturing leverages simulation-based decision support across a spectrum of decision types.
- Applications can be categorized based on their frequency.
- Real-time and operational decision-making covers immediate decisions and near-time operational decisions, like planning machine maintenance in the next shifts.
- Tactical and strategic decision-making is more concerned with long-term planning and design decisions that are less frequent and more permanent.
- Decisions can also be distinguished by the resources involved. At the lowest level, there are the tools or tool sets within a semiconductor wafer fab. At a higher level, the focus shifts to the entire production line, the fab, or material handling systems that connect different tools or tool sets. One step higher still, the internal supply chain and the end-to-end supply chain come into consideration.
- Different tools are used based on the type of decisions.

### Simulation Types

- Simulation-based decision-making approaches are often distinguished by the level of integration between the simulation and the corresponding physical entity. At the one extreme, we can find offline simulations that are solely used to plan and design. At the other extreme, we can find fully integrated digital twins that enable a real-time connection from the simulation model to the physical entity and vice versa.
- While online models can be used concurrently with real-world processes, offline models provide a way to plan and analyze the operations through simulation. These models provide a valuable tool to validate models in the complex semiconductor manufacturing environment.

- Furthermore, with the increased usage and interest in machine learning to optimize those models, simulations can be used for training or to generate data for training.
- Last, simulation results can be extracted and used in a production environment. This approach is limited by the computational power needed, which makes it necessary to either limit the simulation complexity, not use edge computing or only apply it for decisions with longer time horizons.

### Challenges in Industry and Academia

The Dagstuhl Seminar highlighted several critical challenges in this field for industry practitioners as well as researchers.
- Sharing data between operations and simulation models is complicated due to the specific requirements of semiconductor fabrication, which involves highly specialized machines and complex processes spanning sometimes over a thousand processing steps and more than ten weeks.
- Due to cybersecurity concerns of cloud computing, wafer fabs often only use local servers for controlling the production environment.
- The restriction to local infrastructure makes it harder to collect and process the data for simulations, as well as feed in results from simulation-based decision support into the system.
- Integrating simulation-based decision support into the system is therefore a substantial task.
- Another significant industry challenge pertains to the integration of machine learning into simulation-based decision support. Addressing how to effectively employ machine learning, particularly reinforcement learning, in semiconductor manufacturing emerges as a critical concern.
- Moreover, integrating operational measures to achieve business objectives is a difficult challenge due to the complex environment. Measuring the impact of operational changes in the sophisticated semiconductor manufacturing process and aligning them with business goals presents unique difficulties, although simulation has proven to be a valuable tool in this regard.
- Finally, a recurrent issue is the reusability of simulations and models. The intricate manufacturing environment makes it necessary to simulate only parts of the manufacturing process with different levels of detail. This leads to rework due to the necessity of creating different simulation models for the same physical entities.
- Additionally, the proposal of a standardized language with a digital ontology model as a reference point aims to enhance reusability.

In academia, the challenges are shaped by the ever-evolving landscape of semiconductor manufacturing.
- Testbeds for simulations, comprising benchmarking data and industrial and simulation data, are deemed essential for addressing the most pressing challenges faced by researchers.
- Finally, simulation optimization emerged as a key topic of discussion, emphasizing the importance of developing and implementing effective simulation models for achieving optimal outcomes within semiconductor manufacturing.

### Conclusions

- Simulation-based decision support plays a crucial role in semiconductor manufacturing optimization.
- Its applications span from real-time operational decisions to long-term strategic planning, encompassing various levels of the manufacturing process.
- Simulation types vary from offline models for planning and design to fully integrated digital twins for real-time control.
- However, the industry faces challenges in data sharing, cybersecurity, machine learning integration, operational alignment with business goals, and simulation reusability. Addressing these challenges is vital for semiconductor manufacturers to enhance their operations and stay competitive in this rapidly evolving industry.
- In academia, the emphasis lies on testbeds, simulation optimization, and fostering active collaboration with the industry to acquire essential insights and advance research in semiconductor manufacturing simulation.

## 5.5 Breakout Session 2b

*Michael Hassoun (Ariel University, IL), William Bitsch (WHU – Vallendar, DE), Stéphane Dauzère-Pérès (Mines Saint-Etienne, FR), Hans Ehm (Infineon Technologies – München, DE), Rohan Korde (Arizona State University – Tempe, US), Giulia Pedrielli (Arizona State University – Tempe, US), and Mahsa Shekari (Polytechnic University of Milan, IT)*



### Background and semiconductor manufacturing simulation major characteristics

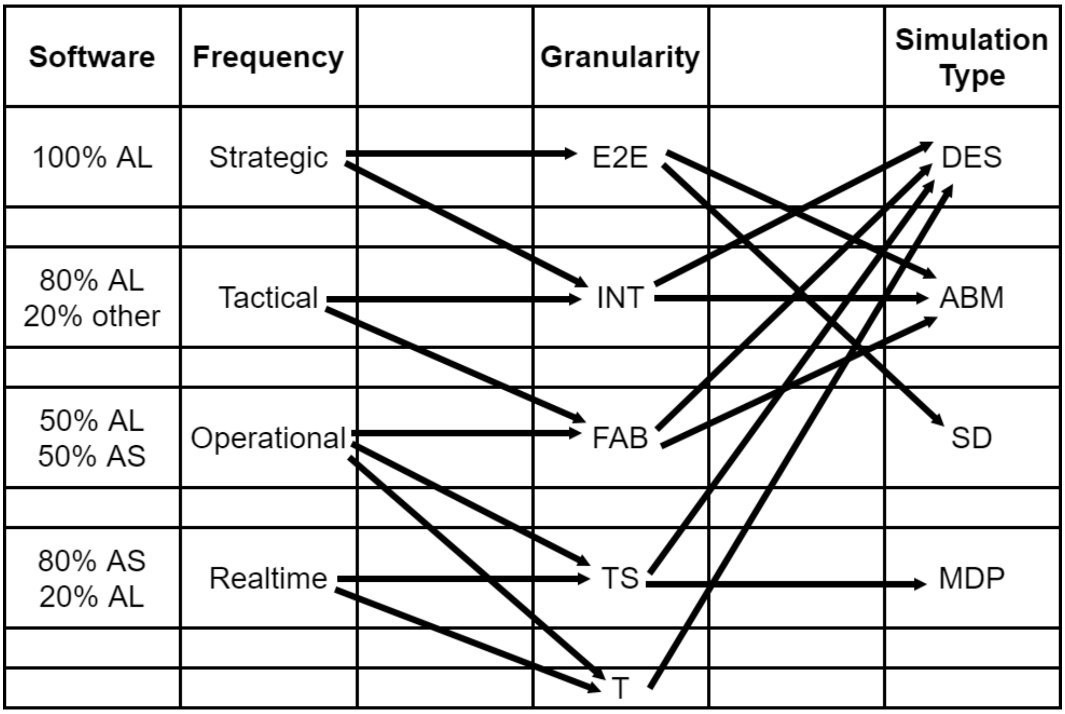
### Granularity

In general, simulation in semiconductor manufacturing is carried out at these levels:
- End to end supply chain (E2E). From raw material suppliers to end customers
- Internal supply chain (INT). Fabrication, Sort, Assembly/Test
- Factory (FAB). The fabrication itself. Usually withing a single fab, production and transport.
- Tool set (TS). Usually the external operation of a group of identical tools, production and transport.
- Tool (T). Usually the internal sequence of operations inside the machine itself.

### Types of simulations

- Discrete event simulation (DES)
- Agent-based modeling (ABM)
- System dynamics (SD)
- Differential equations and Markov Decision Process (MDP)

**Table 3** Connection Between the Different Simulation Dimensions.




### Frequency of use/ type of problem

- Strategic (Long term capacity planning)
- Tactical (Medium term capacity planning, specific mix scenarios, capacity adjustments etc.)
- Operational (Day or shift planning. Actual WIP situation)
- Real Time: (Immediate decision-making at the lot/PM/tool level)

### Software

Typically, two simulation software tools are used in semiconductor manufacturing:

- AnyLogic
- Autosched AP

The connections between **SOFTWARE, FREQUENCY, GRANULARITY,** and **SIMULATION TYPE** are captured in Table 3.

### Future needs

Specific areas in the fab that require attention, and simulation is most probably the tool of choice.

We first point out the need to develop simulation to tackle operation questions that are not yet solved. The most prominent examples of such operational topics are, amongst others, reticle management, and preventive maintenance.

Open questions in terms of simulation abilities:

1. The team has recognized that simulation models developed to tackle certain issues are seldom reusable for other inquiries. Lots of rework is required because old models need to be modified to accommodate business-level changes. The question is how can simulation models be designed and documented in such a way to ease their reuse?

2. Semiconductor-specific simulation toolbox: Today, only Autosched AP offers a semiconductor oriented set of tools and functionalities. There seem to be a need for such a dedicated semiconductor-specific simulation toolbox similar for example to the transportation toolbox that can be found in AnyLogic
3. How to create a standardized language with an ontology for simulation? Create a "role model" as a digital reference (NXP, STMicro, Infineon).
4. How to create fab-level testbeds for simulation models? Several models exist, the latest to have been presented to the community as a replacement for the aging MIMAC is the SMT2020. Yet, their complexity makes them difficult to implement. Furthermore, the data format, although "open" (Excel tables) fits the Autosched software which is expensive to purchase. As a result, University struggles to implement such testbeds for research purposes. An open-source tool, already implementing the baseline model while allowing customization would greatly ease the use of such testbeds.
5. How to capture industrial data in the testbeds? Actual industrial data (WIP/availability/qualification/etc. snapshots) has been made available to researchers. How can such data be integrated to testbed to test, for example, scheduling policies?
6. How and when to integrate simulation with optimization, and what is the adequate level of detail of simulation model inside an optimization model?
7. How can digital twins be integrated with simulation models to manage operations more effectively?
8. How can machine learning models be integrated with simulation models to manage operations more effectively?

## 5.6 Breakout Session 3a

*Lars Mönch (FernUniversität in Hagen, DE), William Bitsch (WHU – Vallendar, DE), Stéphane Dauzère-Pérès (Mines Saint-Etienne, FR), Michael Hassoun (Ariel University, IL), and Daniel Schorn (FernUniversität in Hagen, DE)*



### Sustainability (in a narrow sense)

- Wafer fabs highly energy-intensive
- annual energy utility bills huge
- emission of non-CO2 greenhouse gases, such as perflorocarbons (PFCs), extremely long atmospheric lifetime huge

### Topics

- energy reduction vs. energy management
- pollution reduction (CO2/greenhouse gases)
- Waste management => "control" wafers, nonproductive (NP) wafers
- Water management (water needs to be cleaned, reuse rate)
- Introduction of renewable energy sources (such as photovoltatics (PVs), wind turbines (WTs)), considering pollution taxes and/or subsidies => strategic network design problems
- Workforce: work from home

### Scheduling

- Clean room requires around 60% of overall energy consumption
- Management of peak demand (during summer time)
- Maybe backend more appropriate compared to frontend

### Conclusion

- Getting data (energy consumption by tools, energy offered by PV, WT) challenging
- Which topics can be addressed by OR tools?
- It is important to generate/store energy on the fab level?
- A Survey is needed, not much work done so fare.

### Breakout Session 3b

The discussions in this session were not collected.

### Breakout Session 3c

The discussions in this session were not collected.

## Participants


- Dominik Bisslich
Infineon Technologies AG – Neubiberg, DE
- William Bitsch
WHU – Vallendar, DE
- Stéphane Dauzère-Pérès
Mines Saint-Etienne, FR
- Hans Ehm
Infineon Technologies – München, DE
- John Fowler
Arizona State University – Tempe, US
- Michael Hassoun
Ariel University, IL
- Jessica Hautz
KAI – Villach, AT
- Cathal Heavey
University of Limerick, IE
- Raphael Herding
Westfälische Hochschule – Bocholt, DE
- Young Jae Jang
KAIST – Daejeon, KR
- Adar Kalir
Intel Israel – Qiriat-Gat, IL
- Rohan Korde
Arizona State University – Tempe, US
- Peter Lendermann
D-SIMLAB – Singapore, SG
- Andrea Matta
Polytechnic University of Milan, IT
- Lars Mönch
FernUniversität in Hagen, DE
- Giulia Pedrielli
Arizona State University – Tempe, US
- Thomas Ponsignon
Infineon Technologies – München, DE
- Alexandru Prisacaru
Bosch GmbH – Stuttgart, DE
- Oliver Rose
Universität der Bundeswehr – München, DE
- Henrik Schmielau
Infineon Technologies – München, DE
- Daniel Schorn
FernUniversität in Hagen, DE
- Masha Shekari
Polytechnic University of Milan, IT
- Liji Shen
WHU – Vallendar, DE
- Marcel Stehli
Globalfoundries – Dresden, DE
- Gian Antonio Susto
University of Padova, IT


Report from Dagstuhl Seminar 23371

# Roadmap for Responsible Robotics

**Michael Fisher**[*1], **Marija Slavkovik**[*2], **Anna Dobrosovestnova**[*3], and **Nick Schuster**[*4]

1 **University of Manchester, GB.** michael.fisher@manchester.ac.uk
2 **University of Bergen, NO.** marija.slavkovik@uib.no
3 **TU Wien, AT.** anna.dobrosovestnova@tuwien.ac.at
4 **Australian National University – Canberra, AU.** nick.j.schuster@gmail.com

**Abstract**

This report documents the program and the outcomes of Dagstuhl Seminar 23371 "Roadmap for Responsible Robotics". The seminar was concerned with robots across all their forms, particularly autonomous robots capable of making their own decisions and taking their own actions without direct human oversight. The seminar brought together experts in computer science, robotics, engineering, philosophy, cognitive science, human-robot interactions, as well as representatives of the industry, with the aim of contributing to the steps towards ethical and responsible robotic systems as initiated by actors such as the European Robotics Research Network (EURON), the European Union's REELER, and others. We discussed topics including: "Why do autonomous robots warrant distinct normative considerations?", "Which stakeholders are, or should be, involved in the development and deployment of robotic systems, and how do we configure their responsibilities?", "What are the principal tenets of responsible robotics beyond commonly associated themes, namely trust, fairness, predictability and understandability?". Through intensive discussions of these and other related questions, motivated by the various values at stake as robotic systems become increasingly present and impactful in human life, this interdisciplinary group identified a set of interrelated priorities to guide future research and regulatory efforts. The resulting roadmap aimed to ensure that robotic systems co-evolve with human societies so as to advance, rather than undermine, human agency and humane values.

**Seminar** September 10–15, 2023 – https://www.dagstuhl.de/23371

**2012 ACM Subject Classification** Computer systems organization → Robotic autonomy; Social and professional topics → Codes of ethics; Security and privacy → Trust frameworks; Software and its engineering → Software verification and validation

**Keywords and phrases** Robotics, Responsibility, Trust, Fairness, Predictability, Understandability, Ethics

**Digital Object Identifier** 10.4230/DagRep.13.9.103

## 1 Executive Summary

*Michael Fisher*
*Marija Slavkovik*
*Anna Dobrosovestnova*
*Nick Schuster*



The ISO 8373 standard ("Robots and Robotic Devices – Vocabulary") defines a robot as "an actuated mechanism programmable in two or more axes moving within its environment, to perform intended tasks". Aligned with this definition, we consider "robotics" to cover a wide

[*] Editor / Organizer

range of devices – e.g. vehicles, probes, drones, industrial devices, and personal robots – as well as the complex sociotechnical processes surrounding the development and deployment of such systems. Given that robotic systems are increasingly capable of acting without direct human oversight, and that they're being deployed in an increasing variety of contexts, a range of concerns beyond technical reliability emerge. Many authors, across a variety of disciplines, have pointed to the need for "responsibility" in robotic systems. However, while it is popular to highlight this as a target, there is no agreed route to achieving *responsible robotics*. In addition, there is sometimes even little agreement on what *responsibility* here comprises.

The aim of this Dagstuhl Seminar was to identify the key components of responsibility in this context and then, crucially, provide a roadmap for achieving responsible robotics in practice. By doing so, the seminar contributed to the ongoing efforts established with the Roboethics Roadmap put forth in January 2007 by the European Robotics Research Network (EURON), the European Union's REELER, SIENNA, and TECHETHOS projects, and the UK's RoboTIPS project, among others.

In the original proposal of the seminar, four themes commonly associated with responsible robotics were emphasized: *trust*, *fairness*, *reliability*, and *understandability*. In the course of the seminar, however, the participants – comprising philosophers, engineers, roboticists, cognitive scientists, and industry representatives – identified a broader range of concerns. Firstly, some discussions focused on what responsibility means from different disciplinary perspectives and how these apply to the development, deployment, use, and disposal of robots. In these discussions, it was emphasized that the very term "responsibility" is ambiguous in philosophy and law. The ambiguity and the complexity of the term is, however, rarely reflected in the debates on responsibility in the context of AI and robotics. Referring to [1], responsibility gaps in sociotechnical systems were discussed. We converged on an understanding of responsible robotics as broadly capturing the idea that various parties involved in development, deployment, integration, and maintenance of robots need to be acting in a responsible manner. This involves behaving ethically in their various roles, building ethically sensitive robots, and ultimately taking responsibility for how robotics as a field progresses and how robots are used. This includes "role responsibility", relating to specific functions in robotics; "professional responsibility", which covers obligations in the robotics profession; "moral responsibility", involving ethical decision-making and anticipation of consequences; "legal responsibility", pertaining to compliance with relevant laws and regulations; "social responsibility," regarding the broader impacts of robotic systems on human societies; and "environmental responsibility," regarding their impacts on the natural environment.

As an important step to ensure responsible robotics, discussions considered the diverse roles and responsibilities of key stakeholders, including businesses, universities, governments, users, and others who stand to affect, or be affected by, robotic systems. Specifically, it was noted that universities play a crucial role in shaping the professionals who design, engineer, and operate robotic systems. Engineering and design curricula should thus include modules on responsible innovation, safety standards, and the potential consequences of misuse. This could be done by intensifying the dialogue and collaborations with other disciplines, in particular humanities and social sciences, following promising initiatives such as Embedded EthiCS. To align robotics with ethical standards, businesses in turn must conduct thorough risk assessments, addressing potential misuses and implementing safeguards in their products. For example, in the case of AI-based robotic systems, providers may rely on existing risk management frameworks such as the one recently developed by the

National Institute of Standards and Technology for AI system (`https://www.nist.gov/itl/ai-risk-management-framework`). Additionally, they should provide comprehensive user manuals, conduct user training programs, and actively collaborate with regulatory bodies to establish industry-wide standards. Transparent communication about the capabilities and limitations of their products is essential to ensure that users have a clear understanding of how to responsibly engage with robotic technologies. Furthermore, governments play a pivotal role in creating and enforcing regulations that govern the use of robotic products and services. They must collaborate with industry experts to establish ethical guidelines, safety standards, and legal frameworks. Regulatory bodies should continuously update these frameworks to keep pace with technological advancements. Furthermore, governments should invest in public awareness campaigns to educate citizens about the benefits and risks of robots, mitigating the potential for misuse or misunderstanding.

Discussions also emphasized that an extended definition of responsibility, encompassing not only technical but also social and political considerations, requires a similarly expansive understanding of trust, fairness, reliability, and understandability as well as the addition of other normative concepts. To address this, other potentially relevant concepts were identified through an iterative voting exercise. The final list included: *dignity*, the inherent worth of each member of the moral community who stands to be impacted by robotic systems; *autonomy*, enabling human beings to act in accordance with their own interests and aspirations; *privacy*, empowering people to protect and share sensitive information about themselves as they see fit; *safety*, protecting the various aspects of physical and emotional well-being; *trust*, ensuring that people have good reason to believe that robotic systems are aligned with their legitimate interests; *justice/fairness,* making the impacts of robotic systems acceptable to all who stand to be affected by them; *accountability,* ensuring that the right agents are held to account for adverse outcomes; and *sustainability,* regarding the impacts of robotic systems on the natural world and future generations. It was not our objective to generate an exhaustive list. Rather, the list reflected the principle concerns that emerged from discussion of current and near-future uses and capabilities of robotic systems.

In summary, apart from the group level discussions, 4 working groups were held: These included working groups on:

- Fairness
- Trust
- Why robots require different considerations?
- Predictability

In sum, the main outcome of the seminar was a draft of a document developed collectively and encompassing these and other related topics. The document is intended for a wide range of stakeholders and relevant, affected parties, including researchers, policymakers, industry leaders, practitioners, NGOs, and civil society groups. Recognizing that the current group of authors primarily represents research perspectives (and those coming primarily from the Global North), we are aware of the necessity to incorporate a broader array of viewpoints. Therefore, we are committed to including more diverse perspectives in this discussion going forward, to inform future versions of this roadmap, to better promote the development of responsible robotics, and to help navigate the complex sociotechnical terrain that lies ahead.

## References

1 Santoni de Sio, Filippo, and Giulio Mecacci. Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, 34: 1057-1084, 2021.

## 2 Table of Contents

# 3 Overview of Talks

## 3.1 Fairness in robotics: a philosophical approach

*Helen Beebee (University of Leeds, GB)*



"Fairness" is a concept that crops up a lot in discussions of AI ethics. However it is, I claim, often used in a much broader sense than in philosophy. In philosophy fairness is a concept that is normally deployed in the context of "distributive justice" – that is, where what is at issue is the "fairest", or most just, allocation of resources, opportunities, and so on across society. Fairness, however, is generally not carefully defined. A straightforward and relatively uncontroversial definition might be: A process or rule – and, by extension, an outcome or action resulting from it – is fair if it treats everyone equally, unless unequal treatment is merited/justified/deserved.

Fairness is specifically about treating people equally, and not about treating people as they deserve to be treated – which is a much broader concept. (Nobody deserves to be burgled, but the fact that X was burgled while other people were not is not inherently unfair.)

In AI, fairness seems to have been singled out as a distinctively important moral concept. This is perhaps merited when it comes to the kinds of bias that can arise in decision-making based on the application of large-scale demographic data to a particular case, as is sometimes the case in machine learning. This does apply to come extent to robotics, and – when it does apply – the relevant considerations are how a given demographic generalisation is being used, whether there is differential treatment, and, if there is, whether that differential treatment is justified. But behaving fairly is just one way of behaving well. Robots operate in local situations, and – just like humans – they need to behave well more generally, and not just fairly. So it is not at all clear that "fairness" is more important than other moral concepts when it comes to robotics.

## 3.2 Responsibility in Autonomous Systems

*Michael Fisher (University of Manchester, GB)*


**Main reference** Louise A. Dennis, Michael Fisher: "Verifiable Autonomous Systems: Using Rational Agents to Provide Assurance about Decisions Made by Machines", Cambridge University Press, 2023.
**URL** https://doi.org/10.1017/9781108755023

Responsibility in Robotics comprises two, related aspects:
1. The responsible development of robotic systems; and
2. The responsibility that our robots have, especially once they become autonomous.

Concerning (1) there is already a vast literature on "responsible innovation" and, while we must build on that, we need also take into account the issues relevant to robotics and particularly autonomous robotics. Work on standards in these areas is important, for example the British Standards Institution "Guide to the Ethical Design and Application of Robots and Robotic Systems" (BS8611), published in 2016 and revised in 2023, as well as related work on "Sustainable Robotics" (BS8622). A key aspect here is that strong verification techniques (beyond probabilistic estimates) should be required , especially where robots are to be involved in critical issues. In all this influencing institutions/government/regulators, etc, is vital. To change/update regulatory guidelines, to stimulate changes in government policies, and to change the way robotic systems are developed.

Once we delegate sufficient agency to a robotic system, making it autonomous, then the decisions the robot might make become crucial. Here, the trustworthiness of autonomous robots becomes central. Although trustworthiness in standard systems often equates to "reliability", the move to more autonomous systems expands trustworthiness so that it must incorporate beneficiality – that we believe the robot is making its decisions for our benefit. Such views support a move to more nuanced architectures (e.g. neuro-symbolic) providing better ways to build autonomous robots and making predictability, understandability, fairness, and trustworthiness easier

**References**


1 British Standards Institution. BS 8611: Guide to the Ethical Design and Application of Robots and Robotic Systems. `https://standardsdevelopment.bsigroup.com/projects/9021-05777`. Revised 2023.

2 Raja Chatila, Virginia Dignum, Michael Fisher, Fosca Giannotti, Katharina Morik„ Stuart Russell, and Karen Yeung. *Trustworthy AI.* In *Reflections on Artificial Intelligence for Humanity*, pages 13–39. Springer, 2021.

3 Louise A. Dennis and Michael Fisher. *Verifiable Autonomous Systems: Using Rational Agents to Provide Assurance about Decisions Made by Machines.* Cambridge University Press, 2023.

4 Michael Fisher, Viviana Mascardi, Kristin Yvonne Rozier, Bernd-Holger Schlingloff, Michael Winikoff, and Neil Yorke-Smith. Towards a Framework for Certification of reliable autonomous systems. *Autonomous Agents and Multi Agent Systems*, 35(1):8, 2021.

5 M. R. Mousavi, A. Cavalcanti, M. Fisher, L. Dennis, R. Hierons, B. Kaddouh, E. L.-C. Law, R. Richardson, J. O. Ringer, I. Tyukin, and J. Woodcock. Trustworthy Autonomous Systems Through Verifiability. *Computer*, 56(2):40–47, 2023.

6 Interest Group on Neuro-Symbolic AI. Alan Turing Institute. `https://www.turing.ac.uk/research/interest-groups/neuro-symbolic-ai`.


## 3.3 The Importance for Robots of Knowing When They Don't Know

*Michael Milford (Queensland University of Technology – Brisbane, AU)*


**Joint work of** Helen Carson, Jason Ford and Michael Milford
**Main reference** Helen Carson, Jason J. Ford, Michael Milford: "Unsupervised Quality Prediction for Improved Single-Frame and Weighted Sequential Visual Place Recognition", in Proc. of the IEEE International Conference on Robotics and Automation, ICRA 2023, London, UK, May 29 – June 2, 2023, pp. 3955–3961, IEEE, 2023.
**URL** https://doi.org//10.1109/ICRA48891.2023.10160679

As robotics and automation both matures and stalls due to hard deployment challenges, it looks increasingly likely that fully automated, unsupervised systems will only make up a subset of total robot deployments, for both capability and operational concerns. Instead much robotic deployment will occur in collaborative and semi-supervised environments, where robots and people both play a role and interact in rich and meaningful manners beyond simple supervision and oversight. To maximize both the capability of these autonomous systems as well as their collaborative potential with human operators, both present and remote, these autonomous systems will need to "know when they don"t know" – the power of introspection. Introspection enables graceful performance degradation, but also facilitates handover to human operators. From a pragmatic point of view, for any safety or operationally critical activity, a system with a certain level of performance, say in terms of accuracy, but no introspection capability will often be vastly inferior to a system with slightly lower accuracy but good introspection capability. Whilst introspection and related concepts like verification are mature, well practiced areas in domains like aerospace, their consideration and treatment in robotics is relatively early stage or not done. We propose that a substantial research investment and focus in introspection for robotics and autonomous systems will pay dividends, both in terms of advancing knowledge but particularly in enabling promising robotic technologies to successfully make the transition into trusted, enduringly deployed systems.

## 3.4 Trust and Interactive Robotics

*AJung Moon (McGill University – Montreal, CA)*



This talk unpacks the many ways in which the word "trust" is and has been used in robotics, human-robot interaction, and AI ethics. Between "trustworthy AI" and user trust in specific capabilities of a robot, there is a large gap and diversity of solutions to address the trust-trustworthiness problem.

Building on the current trends in AI ethics (namely, principle-based approaches toward trustworthy AI and framing of ethics issues as model/product-level fairness/transparency/accountability problems) and the temptation by roboticists to borrow much of AI ethics contents directly for "responsible robotics," I problematize how these trends can fail us in our attempts to build generally good robotic systems.

I describe this as a 'water bottle model of trust'. I argue that responsible robotics is an exercise that should help build trustworthy robotics design norms that focus on considerate forms of design and deployment of all robots, rather than one that narrowly guides the ethical design of a single system/hardware/feature. In this process, we should challenge our existing assumptions/taboos (e.g., those related to anthropomorphization) and think about what our shared vision of the world with robots looks like.

## 3.5 Encouraging Inferable Behavior for Autonomy: Repeated Bimatrix Stackelberg Games with Observations

*Ufuk Topcu (University of Texas – Austin, US)*


**Joint work of** Mustafa O. Karabag, Sophia Smith, David Fridovich-Keil, Ufuk Topcu

When interacting with other non-competitive decision-making agents, it is critical for an autonomous agent to have inferable behavior: Their actions must convey their intention and strategy. For example, an autonomous car's strategy must be inferable by the pedestrians interacting with the car. We model the inferability problem using a repeated bimatrix Stackelberg game with observations where a leader and a follower repeatedly interact. During the interactions, the leader uses a fixed, potentially mixed strategy. The follower, on the other hand, does not know the leader's strategy and dynamically reacts based on observations that are the leader's previous actions. In the setting with observations, the leader may suffer from an inferability loss, i.e, the performance compared to the setting where the follower has perfect information of the leader's strategy. We show that the inferability loss is upper-bounded by a function of the number of interactions and the stochasticity level of the leader's strategy, encouraging the use of inferable strategies with lower stochasticity levels. As a converse result, we also provide a game where the required number of interactions is lower bounded by a function of the desired inferability loss.

# 4 Working groups

## 4.1 Are robots different from AI? Definition and (some) Related Considerations

*Anna Dobrosovestnova (TU Wien, AT)*

**Joint work of** Kevin Baum, Anna Dobrosovestnova, AJung Moon, SeverinLemaignan, Pericle Salvini, Sara Ljungblad, Kate Devitt Ufuk Topcu, Raja Chatila, Teresa Scantamburlo


According to the definition of robot laid out in in the ISO 8373:2012 standard (International Organization for Standardization [ISO], 2012)): robot is an actuated mechanism programmable in two or more axes with a degree of autonomy (i.e., the ability to perform intended tasks based on current state and sensing, without human intervention), moving within its environment, to perform intended tasks. This definition allows us to distinguish between robots and other automated systems. Specifically, it implies that a robot is, first and foremost, a physical piece of machinery. This already excludes software, e.g. software bots, voice assistants, or image recognition from the broad category of robots. Furthermore, the definition underscores how robots require some degree of autonomy. Sostero (2020) points out certain ambivalence when it comes to anchoring what autonomy means because current state of the art technology and existing regulations allow only for limited autonomy. That said, this means the given definition of robot also excludes mobile machinery that only follows pre-programmed instructions without coupling between the machinery and the environment (e.g. 3D printers). To summarize, robots can be considered intelligent embodied agents situated in the real world, which means their existence and operation occur in the real world.

While many concerns related to ethical implications and responsibility overlap between robots and AI systems, the embodied and (partly) autonomous nature of robots bring with it a host of considerations relevant in the context of the broader conversation about ethics of robotics and responsibility. Firstly, the physicality of the robotic systems mean they can cause physical harm and cause injury. Secondly, the embodied nature of robots, coupled with autonomous movement and perceived goal orientedness is known to elicit in people a tendency to respond and treat robots as (quasi-)social actors. This tendency is further enabled by the fact that the so called social robots are increasingly developed to look and behave like humans. The potential dangers of designed and/or in perceived robot sociality have already been discussed in the robot ethics and related literature in relation to deception, unilateral bond, and how such robots can reshape affect and relationality laden practices e.g., when it comes to robots deployment in service sectors. Beyond these issues, we also identified a host of challenges related to the uncertainty of deployment the robots in the physical world. For instance, the robot physicality also implies that we cannot simply translate design assumptions and practices borrowed from the software industry. For example, deciding to terminate interactions with, or use of a robot, is a different process when compared to uninstalling a software or a mobile app. Turning off a robot does not mean the robot no longer impacts the spaces wherein it is present. This can have various implications, ranging from concerns for data privacy to material sustainability. Likewise, some of the robots provide crucial physical assistance to those who have various impairments. Ceasing service of such system can mean the difference between a person"s ability to conduct activities of daily life by themselves and not.

Based on these, and other concerns stemming from the (physically) embodied and (partly) autonomous character of robots mean, we argue, robots extend the scope of (ethical) and responsibility related considerations beyond what has been addressed in the discourses about ethical and responsible AI.

### References

1 Sostero, M. (2020). Automation and Robots in Services: Review of Data and Taxonomy (JRC Working Papers Series on Labour, Education and Technology 2020/14). European Commission, Joint Research Centre (JRC). `http://hdl.handle.net/10419/231346`.

## 4.2 The Role of Fairness in Responsible Robotics

*Sarah Moth-Lund Christensen*


**Joint work of** Kevin Baum, Helen Beebee, Simon Coghlan, Emily Collins, Kate Devitt, Michael Fisher, Masoumeh Mansouri, Sarah Moth-Lund Christensen

In philosophy, fairness has traditionally not been investigated as a moral concept in itself, but instead played second fiddle in relation to political philosophical concerns such as "distributive justice" [1]. In other words "fairness" has been utilised as a concept regarding the allocation of resources, opportunities, and so on across society. As such, it is not a concept easily and clearly defined in the literature. However, in recent years "Fairness" has been heavily deployed as a key term not as part of theoretical discussions on distributive justice, but instead in opposition to the rising concern regarding bias issues in machine learning models [2].

This working group aimed to broaden the terms of the "Fairness" debate, recognising that "Fairness" as a concept is not and should not be reduced to "Algorithmic Fairness". As such, the group set out to investigate what further role "Fairness" might play with regards to responsible robotics, and as such whether the concept of "Fairness" may be used to identify and illuminate issues regarding contemporary robotics design and deployment. The following inter-related points were identified as potential Fairness concerns relevant to Responsible Robotics design and deployment practices:

Algorithmic Fairness: Algorithmic injustice is an ongoing concern regarding the use of machine learning models. As machine learning can also be used in robots, the ongoing algorithmic injustice debate still relevant to discussion of fairness in relation to robotics.

Fairness and Design: Fairness may play a role even on the lowest level of design. Certain components may have poorer performance for certain demographics, while design choices such as language used can affect for whom the technology is useful for.

Fairness and Accessibility: On a larger scale, the deployment of robots in public and private domains can give rise to concerns regarding financial inaccessibility. As an example, consider a robot developed as a disability aid or as a teaching tool that is prohibitively expensive for the individuals or public schools that are in need of it. As such, societal or systemic structures can give rise to fairness concerns for the deployment of robots.

For completeness, it should be noted that the work group participants throughout the session discussed and expressed concerns on the merit of Fairness as a focal key concept in the Responsible Robotics debate, particularly in comparison to either more established political philosophical concepts such as Justice, or in comparison to other motivating concepts of well-established moral nature. Hence, a conceptual issue emerged regarding the notion of "Fairness", as the lack of current in-depth definitions results in ambiguity and lack of clarity when attempting to use "Fairness" as a guiding principle. As such, further formal investigation in the definition and dimensions of "Fairness" is needed in order to fully establish its full role in Responsible Robotics.

**References**

1 John Rawls. *A Theory of Justice.* Harvard, MA: Harvard University Press, 1971.

2 Reuben Binns. Fairness in Machine Learning: Lessons from Political Philosophy *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, in Proceedings of Machine Learning Research. 81: 149–159., 2018.

## 4.3 Trust and Responsible Robotics

*Nick Schuster (Australian National University – Canberra, AU), Hein Duijf (LMU München, DE), Nadin Kokciyan (University of Edinburgh, GB), and Thomas Michael Powers (University of Delaware – Newark, US)*



Responsible robotics requires that robots, as well as the people and organizations who design and deploy them, are trustworthy. This goes beyond getting people to trust robotic systems, including their human elements, which might be accomplished whether or not people have good reason for trust (e.g. through effective advertising). Rather, robotic systems must satisfy certain independent normative standards in order to warrant trust. What are these standards, and what challenges must robotic systems overcome to satisfy them? We think that trust is importantly distinct from transparency, explainability, and predictability. While these qualities can make robotic systems trustworthy, trust seems especially important where robotic systems aren't, or can't be, made fully transparent, explainable, and predictable. Like human-human interactions, human-robot interactions can involve unavoidable uncertainty. Also like human-human interactions, human-robot interactions can take place in physical space (as opposed to cyberspace) and therefore make immediate physical harm a real and sometimes visceral possibility. These factors necessitate a trust that's structurally similar to trust between human actors: people often have to trust each other to behave appropriately without full knowledge of each others' motives, intentions, capacities, propensities, needs, vulnerabilities, etc. But robots are different from humans in relevant respects too. For instance, they don't share humans' basic interests in avoiding pain, injury, and death; they can't communicate with humans as humans can with each other; and they can be controlled by humans in ways that humans can't be controlled by each other. These factors pose distinct challenges for trust in robotic systems. A promising approach to clarifying and addressing such challenges would draw on social epistemology and moral psychology, to better understand what undergirds trust(worthiness) in general, as well as engineering and organizational studies, especially human factors, to explore possibilities for designing and deploying robotic systems such that people have good reason to trust them despite, or perhaps even because of, their peculiarities.

## 4.4 Predictability

*Michael Milford (Queensland University of Technology – Brisbane, AU)*

**Joint work of** Michael Milford, Nico Hochgeschwender, Dejanira Araiza-Illan, Alcino Cunha, Andrzej Wasowski, Yi Yang, Francisco J. Rodriguez Lera, Hein Duijf, Raja Chatila, Martin Magnusson


There has been some work on defining predictability in the context of robotics, and exploring its connection to other relevant properties, namely understandability. Although not entirely consensual, the common idea of all these definitions is that predictability is about matching the expectations of the user/observer and that predictability lies in a continuum (given a goal, a robot is as predictable as its chosen plan matches the expectations of the user/observer for that goal).

Full predictability might not always be a desirable property for all different users/observers (for a robot operating in public spaces, fully predictable behavior might open opportunities for observers to abuse/bully the robot), so a key responsibility at design time is precisely to identify the level of predictability that is adequate for each stakeholder. Predictability requires the user/observer to know the goal – the design should clarify how this will be achieved, either by designing the robot to also be understandable/legible, building single purpose robots, educating the users (this one a responsibility at deployment time...), etc. Not all users have the same expectations of what is the best plan, so responsible design for predictability should incorporate in the robot some mechanism for the robot to adapt to the individual users, so that (at least) predictability improves over time, or again level everyone by educating at deployment time...

Predictability is also a technical concept: regardless of the user/observer and the robot platform, task and domain, the extent and specificity with which a robot"s actions can be predicted also varies. For example, a large robot moving with substantial inertia through the environment – such as an autonomous truck – has a highly predictable set of next step possibilities – it will continue to move in the current direction at near the current velocity, possibly with the application of acceleration or braking changing its velocity. A human observer does not need to know anything about the algorithms or control systems for the robot in order to have broad predictability for the autonomous truck – it will likely continue on its current trajectory in the next moment, but may increase or decrease its velocity and its heading may change (initially not by very much).

Continuing the autonomous truck example, another key aspect of predictability is predicting the performance of the system. For autonomous vehicles, localization – knowing where the vehicle or robot is located – is a key estimation task that enables safe navigation and higher level behaviours. One aspect of the predictability of a localization system is predictability of how well it is performing – also relating to the concept of introspection. Imagine a choice of two localization systems: one that works well 99 % of the time but is unable to predict its failures that remaining 1% of the time, versus a second system that works well 95% of the time, but is able to predict when it is performing badly 95% of the time. An autonomous vehicle using the first system will unknowingly navigate using incorrect localization information 1% of the time: using the second system, this percentage drops to 0.25%, a major different for such a safety critical application. A side note relates to the research culture of robotics and related fields currently: one key issue with research in this domain currently is that the former system is much more likely to yield a top tier publication, despite the second system having far more utility for many end-user applications.

## Participants

- Dejanira Araiza-Illan
Johnson & Johnson –
Singapore, SG
- Kevin Baum
DFKI – Saarbrücken, DE
- Helen Beebee
University of Leeds, GB
- Raja Chatila
Sorbonne University – Paris, FR
- Sarah Christensen
University of Leeds, GB
- Simon Coghlan
The University of Melbourne, AU
- Emily Collins
University of Manchester, GB
- Alcino Cunha
University of Minho – Braga, PT
& INESC TEC – Porto, PT
- Kate Devitt
Queensland University of
Technology – Brisbane, AU
- Anna Dobrosovestnova
TU Wien, AT
- Hein Duijf
LMU München, DE
- Vanessa Evers
University of Twente –
Enschede, NL
- Michael Fisher
University of Manchester, GB
- Nico Hochgeschwender
Hochschule Bonn-Rhein-Sieg, DE
- Nadin Kokciyan
University of Edinburgh, GB
- Severin Lemaignan
PAL Robotics – Barcelona, ES
- Sara Ljungblad
University of Gothenburg, SE &
Chalmers University of
Technology – Göteborg, SE
- Martin Magnusson
örebro University, SE
- Masoumeh Mansouri
University of Birmingham, GB
- Michael Milford
Queensland University of
Technology – Brisbane, AU
- AJung Moon
McGill University –
Montreal, CA
- Thomas Michael Powers
University of Delaware –
Newark, US
- Daniel Fernando Preciado
Vanegas
Free University of
Amsterdam, NL
- Francisco Javier Rodríguez
Lera
University of León, ES
- Pericle Salvini
EPFL – Lausanne, CH
- Teresa Scantamburlo
University of Venice, IT
- Nick Schuster
Australian National University –
Canberra, AU
- Marija Slavkovik
University of Bergen, NO
- Ufuk Topcu
University of Texas – Austin, US
- Andrzej Wasowski
IT University of
Copenhagen, DK
- Yi Yang
KU Leuven, BE

Report from Dagstuhl Seminar 23372

# Human-Centered Approaches for Provenance in Automated Data Science

**Anamaria Crisan*[1], Lars Kotthoff*[2], Marc Streit*[3], and Kai Xu*[4]**

1 **Tableau Software – Seattle, US**
2 **University of Wyoming – Laramie, US**
3 **Johannes Kepler Universität Linz, AT**
4 **University of Nottingham, GB**

**Abstract**

The scope of automated machine learning (AutoML) technology has extended beyond its initial boundaries of model selection and hyperparameter tuning and towards end-to-end development and refinement of data science pipelines. These advances, both theoretical and realized, make the tools of data science more readily available to domain experts that rely on low- or no-code tooling options to analyze and make sense of their data. To ensure that automated data science technologies are applied both effectively and responsibly, it becomes increasingly urgent to carefully audit the decisions made both automatically and with guidance from humans.

This Dagstuhl Seminar examines human-centered approaches for provenance in automated data science. While prior research concerning provenance and machine learning exists, it does not address the expanded scope of automated approaches and the consequences of applying such techniques at scale to the population of domain experts. In addition, most of the previous works focus on the automated part of this process, leaving a gap on the support for the sensemaking tasks users need to perform, such as selecting the datasets and candidate models and identifying potential causes for poor performance.

The seminar brought together experts from across provenance, information visualization, visual analytics, machine learning, and human-computer interaction to articulate the user challenges posed by AutoML and automated data science, discuss the current state of the art, and propose directions for new research. More specifically, this seminar:

- articulates the state of the art in AutoML and automated data science for supporting the provenance of decision making,
- describes the challenges that data scientists and domain experts face when interfacing with automated approaches to make sense of an automated decision,
- examines the interface between data-centric, model-centric, and user-centric models of provenance and how they interact with automated techniques, and
- encourages exploration of human-centered approaches; for example leveraging visualization.

**Seminar** Spetember 10–15, 2023 – https://www.dagstuhl.de/23372

**2012 ACM Subject Classification** Human-centered computing → Visualization; Information systems → Data provenance; Theory of computation → Mathematical optimization; Human-centered computing → Human computer interaction (HCI); Computing methodologies → Machine learning; Computing methodologies → Search methodologies

**Keywords and phrases** Dagstuhl Seminar, Provenance, AutoML, Data Science, Information Visualisation, Visual Analytics, Machine Learning, Human-Computer Interaction

**Digital Object Identifier** 10.4230/DagRep.13.9.116

* Editor / Organizer



Human-Centered Approaches for Provenance in Automated Data Science, *Dagstuhl Reports*, Vol. 13, Issue 9, pp. 116–136
Editors: Anamaria Crisan, Lars Kotthoff, Marc Streit, and Kai Xu
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

## 1 Executive Summary

*Anamaria Crisan (Tableau Software – Seattle, US)*
*Lars Kotthoff (University of Wyoming – Laramie, US)*
*Marc Streit (Johannes Kepler Universität Linz, AT)*
*Kai Xu (University of Nottingham, GB)*

This Dagstuhl Seminar brings together an interdisciplinary group of researchers and practitioners, spanning Data Science (DS) and Machine Learning (ML), Visualization and Human-Computer Interactions (HCI), and Provenance; to tackle the challenges in automated data science (AutoDS). We specifically focused on ways that methods from human-centered design approaches and provenance can be leveraged to "open up the black box" of AutoDS, introduce greater observability of these methods, and promote human-machine teaming. We observed that there exist many parallel efforts across different disciplines that have yet to be integrated; our seminar brought together these different perspectives as a first step towards producing a general synthesis of methodologies and techniques for advancing AutoDS.

**Primitives for AutoDS and hybrid modes of automation.** Initial implementations of AutoDS tooling were focused on the so-called CASH problem, combining algorithm selection with parameter optimization, which was exclusively limited to the modeling phase of the data science workflow. More recent work has expanded the scope to include tasks pertaining to data preparation, feature engineering, even model deployment and monitoring for concept drift. Within this expanded end-to-end scope for AutoDS, the individual components of the data science pipeline are often referred to as data science primitives; whether those primitives concern work carried out by a human (i.e., selecting a data set for analysis) or a machine (i.e., hyperparameter tuning) depends on the implementation of the system. Discussions on these data science primitives and the scope of the hybrid automation, where humans and automated processes trade-off work, help frame a discussion around provenance and human-centered design.

**Provenance modalities in an end-to-end AutoDS pipeline.** Existing methodologies for provenance in data analysis focus on three related themes: data provenance, computation provenance, and user provenance. These are often studied separately, while they should be explored together in AutoDS to be fully transparent and auditable. It was identified that modalities of capturing data, computation, and user provenance may not always align and there exist few techniques that attempt their integration. Moreover, user provenance can be especially complex to capture and surface, as the thinking and reasoning behind analysis choices and decisions are much more challenging to capture than data science workflow or user interactions. Many open problems and potential solutions were discussed at the seminar and more details are provided in the following sections.

**Visual and interaction techniques for explainable AutoDS (i.e., model-to-human communication).** Data visualization is a powerful medium to help users understand and analyze complex data (in our case the AutoDS provenance), as well as to create opportunities for domain experts and data scientists to interrogate the pipelines themselves. Visual techniques for provenance of AutoDS pipelines exist (i.e., PipelineProfiler, ATMSeer, ModelLineUpper, AutoVizAI, and Visus) but these focus almost exclusively on modeling and do not consider the broader scope of AutoDS primitives. Seminar participants explored the possibilities and utility of visualizing multiple provenance modalities and across AutoDS primitives to achieve this goal.

**Human-centered approaches to data science and analytics (i.e., human-to-model communication).** Seminar participants acknowledged that humans and automated processes must collaborate in AutoDS, and it becomes necessary to explicitly consider the needs of humans to understand and intervene. Human-centered design encapsulates a broad set of methodologies and techniques for designing technology that interfaces with people. Seminar participants advocated for a broader application in human-centered approaches to ML/AI, including mitigating concerns of "black box" algorithms as discussed earlier. A related research challenge identified is to make DS models more "interactive" so user expertise and knowledge can be more easily incorporated, especially for non-technical domain experts. This can happen during the training of a large model through user "steering" to reduce training time, or after deployment with techniques such as "active learning" to continuously improve the module.

## 2 Table of Contents

# 3 Overview of Talks

## 3.1 Overview of Provenance and Visualization

*Kai Xu (University of Nottingham, GB) and Marc Streit (Johannes Kepler University Linz, AT)*



In this 30-minute overview talk, we aim to provide a summary of the state-of-the-art of the research related to provenance and its application in interactive visualization. We started with an introduction of what provenance is and how the concept is used with data analysis and visualization. When going through the latest research. We group the work by the "why" (the goal of provenance analysis), the "what" (what provenance data is needed for the intended goal), and "how", (how to capture and analyze the captured provenance). We conclude the talk with a list of open challenges that are important to the field and need further investigation.

## 3.2 An Introduction to AutoML

*Lars Kotthof (University of Wyoming – Laramie, US)*



Automated machine learning makes state-of-the-art machine learning accessible to people with little to no background in it. Even for machine learning experts, automated methods are helpful to achieve the best performance with relatively little human effort. In this talk, I will give a high-level overview of the problems that automated machine learning solves and how, after a formal definition of the AutoML problem, I will sketch current solution approaches, issues, open challenges, and potential for application of visualization and provenance approaches.

## 3.3 Automating Data Science: Pipe Dream or Reality?

*Anamaria Crisan (Tableau Research – Seattle, US)*



The lack of data scientists but the desire to analyze large data repositories has spurred the development of methods to automate data science. However, in practice, it is complex to orchestrate related human, model, and data processes. Moreover, it becomes difficult to understand how a decision was made and whether this was done by a human or automated process. In this talk, I provide an overview of research motivating the needs and uses of automation. I discuss the existing techniques and tools as well as their limitations. Finally, I discuss the potential of new types of models (specifically LLMs) to further the automation of data science.

## 3.4 Co-Adaptive Analytics and Guidance

*Mennatallah El-Assady (ETH Zürich, CH)*



Mixed-initiative visual data analysis systems rely on a process of co-adaptation where human and AI agents collaboratively perform data-driven problem-solving and decision-making. The co-adaptive process describes the dynamic learning and teaching process these agents are engaged in during their interaction in the mixed-initiative system. In this talk, I give an overview of the state-of-the-art in co-adaptive analysis, highlighting co-adaptive guidance in visual analytics. Structuring the topic further, I present the recent paper on deriving a guidance typology. To illustrate how such theoretical concepts can be put into practice, I present two interactive approaches for topic model refinement that employ different types of guidance: speculative execution and single-objective agents. Furthermore, I demonstrate the Lotse library as a practical framework for co-adaptive guidance implementation. Lastly, I discuss open questions concerning provenance, AutoML, and evaluation.

## 3.5 Exploring Relationships Between Vis/HCI Theory & Provenance

*Leilani Battle (University of Washington – Seattle, US)*



Visualization theory is often developed at a high level, such as in the form of diagrams, taxonomies, or flow charts. However, these artifacts are difficult to implement in visualization tools. By applying taxonomies to provenance data, such as interaction logs, we could better understand how to make visualization theory more practical.

## 3.6 DeepCAVE: A visualization and Analysis Tool for AutoML

*Tanja Tornede (Leibniz University Hannover, DE)*


**Joint work of** Tanja Torned, René Sass, Eddie Bergman, André Biedenkapp, Frank Hutter, Marius Lindauer
**Main reference** René Sass, Eddie Bergman, André Biedenkapp, Frank Hutter, Marius Lindauer: "DeepCAVE: An Interactive Analysis Tool for Automated Machine Learning", arXiv, 2022.
**URL** https://doi.org//10.48550/ARXIV.2206.03493

Visualizing the process of AutoML and its analysis can be done using DeepCAVE. Besides providing a summary of the experimental setup, it offers methods for objective analysis, budget analysis (in multi-fidelity settings), and hyperparameter analysis, This way, the entire interactive framework allows to efficiently generate insights for AutoML problems and brings the human back in the loop.

## 3.7 Provenance Embedding

*Kai Xu (University of Nottingham, GB), Marc Streit (Johannes Kepler Universität Linz, AT)*


**Joint work of** Conny Walchshofer, Andreas Hinterreiter, Kai Xu, Holger Stitz, Marc Streit
**Main reference** Conny Walchshofer, Andreas P. Hinterreiter, Kai Xu, Holger Stitz, Marc Streit: "Provectories: Embedding-Based Analysis of Interaction Provenance Data", IEEE Trans. Vis. Comput. Graph., Vol. 29(12), pp. 4816–4831, 2023.
**URL** https://doi.org//10.1109/TVCG.2021.3135697

In this talk, we propose a research question that may be of interest to seminar participants for discussion during the seminar. The idea is based on a previous work on modeling and visualizing provenance. The main idea is to capture provenance as a vector sequence, which can then be visualized and analyzed using techniques designed for high-dimensional data. The new idea is to take this one step further, following the process similar to training Large Language Models (LLM) such as ChatGPT by masking a step in the provenance and training a model to predict it. The hope is that such a model can have additional 'intelligence' besides predicting the next step, similar to ChatGPT.

## 3.8 Trrack + Persist

*Kiran Gadhave (University of Utah – Salt Lake City, US)*


**Joint work of** Zach Cutler, Kiran Gadhave, Alexander Lex
**Main reference** Zach Cutler, Kiran Gadhave, Alexander Lex: "Trrack: A Library for Provenance-Tracking in Web-Based Visualizations", in Proc. of the 31st IEEE Visualization Conference, IEEE VIS 2020 – Short Papers, Virtual Event, USA, October 25-30, 2020, pp. 116–120, IEEE, 2020.
**URL** https://doi.org//10.1109/VIS47514.2020.00030

Trrack is a provenance tracking library for the web. One of the goals of the library is to be easy to integrate Trrack. Trrack has a hybrid provenance tracking approach which tracks both actions and state. Trrack stores the diffs between the states to optimize storage.

Computational notebooks have a gap between code and visualization. Semantic, layout and temporal gap are the three highlighted in B2 by Wu et al. [1]. B2 proposes queries (e.g. elections) as a bridge between them. We propose using Trrack provenance to bridge the gap. We've done a Jupyter extension which shows examples of this.

### References

1 Yifan Wu, Joseph M. Hellerstein, and Arvind Satyanarayan. B2: Bridging Code and Interactive Visualization in Computational Notebooks. *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 152–165, 2020, doi: 10.1145/3379337.3415851.

## 3.9 Mosaic

*Dominik Moritz (Carnegie Mellon University – Pittsburgh, US)*


**Joint work of** Jeffrey Heer, Dominik Moritz
**Main reference** Jeffrey Heer, Dominik Moritz: "Mosaic: An Architecture for Scalable & Interoperable Data Views", IEEE Trans. Vis. Comput. Graph., Vol. 30(1), pp. 436–446, 2024.
**URL** https://doi.org//10.1109/TVCG.2023.3327189

Mosaic is an extensible framework for linking databases and interactive views. It links charts, tables, inputs, etc. through a coordinator that optimizes queries and creates data cube indices (for fast linked interactions) as tables in the database. Mosaic is very useful for building linked dashboards and in the future we could track provenance to log it in studies or suggest analyses or subspaces of the data to look at.

## 3.10 Understanding How In-Visualization Provenance Can Support Trade-off Analysis

*Mehdi Chakhchoukh (Université Paris-Saclay, FR)*


**Joint work of** Mehdi Chakhchoukh, Nadia Boukhelifa, Anastasia Bezerianos
**Main reference** Mehdi Chakhchoukh, Nadia Boukhelifa, Anastasia Bezerianos: "Understanding How In-Visualization Provenance Can Support Trade-Off Analysis", IEEE Trans. Vis. Comput. Graph., Vol. 29(9), pp. 3758–3774, 2023.
**URL** https://doi.org//10.1109/TVCG.2022.3171074

In domains such as agronomy or manufacturing, experts need to consider trade-offs when making decisions that involve several, often competing, objectives. Such analysis is complex and may be conducted over long periods of time, making it hard to revisit. In this talk we presented some of our results that were published in an IEEE Transactions on Visualisation and Computer Graphics paper: mainly the idea of refining Ragan et al. [1] purposes for provenance with provenance objects that are task-specific. We discussed if such objects could be used to support the design of provenance visualization for autoML tasks. Finally, we presented the challenges encountered when designing provenance views based on our experience from the experiments we ran with agronomy experts with real-world data and applications.

### References

1 Eric D. Ragan, Alex Endert, Jibonananda Sanyal and Jian Chen. Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes. *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 31-40, 31 Jan. 2016, doi: 10.1109/TVCG.2015.2467551.

## 3.11 Data Provenance for Reproducible Research

*Sheeba Samuel (Friedrich Schiller University – Jena, DE)*


**Joint work of** Sheeba Samuel, Daniel Mietchen
**Main reference** Sheeba Samuel, Daniel Mietchen: "Computational reproducibility of Jupyter notebooks from biomedical publications", CoRR, Vol. abs/2209.04308, 2022.
**URL** https://doi.org//10.48550/ARXIV.2209.04308

Reproducible research refers to the idea that scientific results are documented and published in a way that others may verify the findings and build upon them. Data provenance is one of the integral components of reproducible research across domains. We investigate the computational reproducibility aspect of Jupyter notebooks within the context of publications indexed in PubMed Central. Our research endeavors to identify common challenges and best practices, delineate emerging trends and propose potential enhancements to Jupyter-related workflows associated with publications. To bolster the reproducibility of Jupyter-related workflows, we delve into various data provenance approaches and tools. Specifically, we examine the utility of tools such as ProvBook [1] and MLProvLab [2] in capturing and visualizing diverse aspects of provenance information. MLProvLab, in particular, enables granular tracking of information at both the notebook and cell levels, visualizing dependencies between cells and data within a notebook. This functionality is invaluable for data scientists, as it aids in comprehending the cascading effects of changes made to one cell on subsequent cells and, ultimately, on the research results. Furthermore, we revisit the W3C model, PROV-O [3], for representing provenance information, emphasizing the pivotal role of ontologies in modeling such information effectively. Our exploration extends to the ReproduceMe data model, which facilitates the sharing of computational provenance in a machine-readable format, enhancing the accessibility and utility of provenance information. Finally, we address research questions concerning the significance of provenance information and its utilization in machine learning and deep learning pipelines. We underscore the importance of sharing and harnessing collected provenance information to enhance the transparency, reproducibility, and trustworthiness of research outcomes in these domains.

### References

1 Sheeba Samuel and Birgitta König-Ries. ProvBook: Provenance-based Semantic Enrichment of Interactive Notebooks for Reproducibility, 17th International Semantic Web Conference (ISWC) 2018 Demo Track.
2 Dominik Kerzel, Sheeba Samuel, and Birgitta König-Ries. Towards Tracking Provenance from Machine Learning Notebooks. *Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* , 274-281, 2021, doi: 10.5220/0010681400003064.
3 Paolo Missier, Khalid Belhajjame, and James Cheney. The W3C PROV family of specifications for modelling provenance metadata. *Proceedings of the 16th International Conference on Extending Database Technology*, pages 773–776, New York, USA, 2013.

## 3.12 Welcome to Parameter Land – Visual Parameter Space Exploration

*Klaus Eckelt (Johannes Kepler University Linz, AT)*



We introduce an approach to visualize and navigate the hyperparameter space of machine learning pipelines optimized by AutoML methods. Rather than parallel coordinates plots, we envision the hyperparameter space as a dynamic map, allowing users to intuitively explore, optimize, and discover regions of interest–or simply monitor the optimization process. Providing interactive visualizations, like treemaps or LineUp [1], should ultimately enhance the user experience in AutoML leading to higher trust and informed decision-making in machine learning pipelines.

### References

1 Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. LineUp: Visual Analysis of Multi-Attribute Rankings. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2277–2286, 2013, doi: 10.1109/TVCG.2013.173.

# 4 Working Groups

## 4.1 Terminology

*Alex Endert (Georgia Tech – Atlanta, US) Alexander Lex (Utah University – Salt Lake City, US) Alvitta Ottley (Washington University in St Louis, US) Ana Crisan (Tableau Research, Salesforce – Seattle, US) Camelia D. Brumar (Tufts University – Medford, US) Kai Xu (University of Nottingham, GB) Leilani Battle (University of Washington – Seattle, US) Marc Streit (Johannes Kepler University Linz, AT) Menna El-Assady (ETH Zürich, CH) Nadia Boukhelifa (Université Paris-Saclay, FR)*



Embarking on a journey through the multifaceted domain of data provenance, this working group unfolded discussions and explorations, intertwining the conceptual, terminological, and applicative aspects of provenance data. The exploration, situated within the realms of Human-AI interaction and adaptive systems, traversed through a meticulous terminological exploration and an analytical discourse, rooted in a transcript reflecting upon the dual roles of providence data in adaptive systems. The commitment was not just to comprehend but to unfold and apply this understanding in real-world scenarios, especially where the human and artificial intelligence amalgamate.

The terminological exploration unwrapped the concept of "provenance data", probing its constituents, potential applications, and ontological management, with a nuanced focus on existing frameworks like Google's ontologies. The discussions, while illuminating, also underscored the imperative for a cohesive and adaptable understanding of provenance data, which can seamlessly weave through varied applications and domains, ensuring its utility across a spectrum of use-cases, especially within the intricate fabric of Human-AI interactions. A parallel strand navigated through the active and passive roles of providence data within

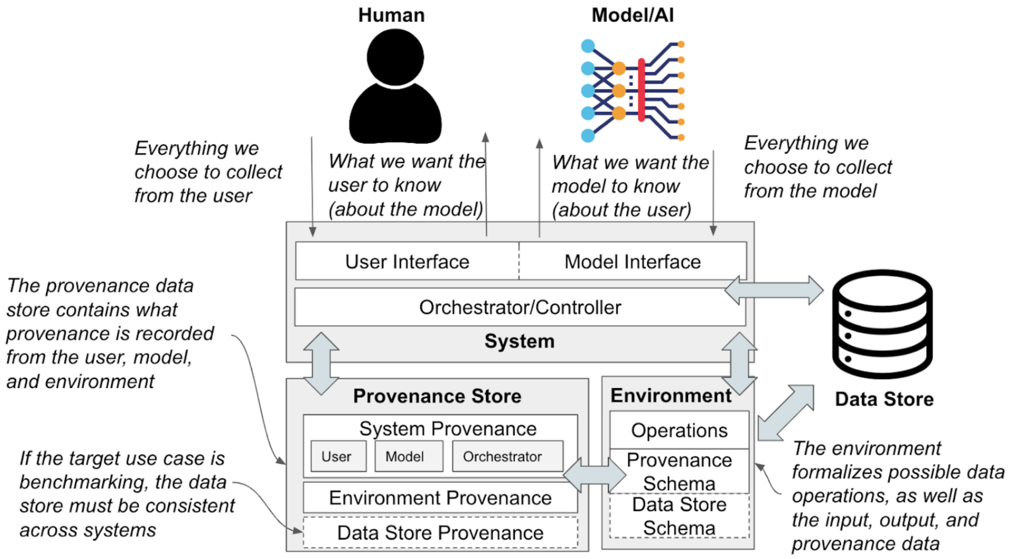



**Figure 1** A conceptual model illustrating the interplay between human actors, artificial intelligence, and a visual analytics system designed to capture provenance data.

adaptive systems and machine learning, spotlighting its potential to not only train models through human interactions but also to elucidate model behaviors, crafting a bi-directional pathway of understanding and application.

Emerging from the discourse was the ontological challenge and a palpable paradox: the need to comprehensively comprehend and define provenance data while concurrently delving into its applications and management in pragmatic contexts. This paradox, particularly pronounced in discussions around adaptive systems, spotlighted the necessity for a clear, structured understanding as pivotal to harnessing providence data's full potential. Simultaneously, it brought forth the challenges and gaps extant in leveraging ontological frameworks across diverse applications and domains, necessitating further exploration and refinement to make these frameworks universally adaptable and coherent.

As the group forges ahead, the commitment is twofold: refining and expanding the understanding and applications of provenance data and ensuring that this theoretical clarity is translatable into pragmatic applications, especially in crafting intuitive, transparent, and effective adaptive systems. This involves not only a deeper exploration and defining of the terminology and conceptual frameworks but also a meticulous examination of its applications, ensuring a seamless transition from theory to practice. Additionally, a continuous, collaborative dialogue with the wider academic and research community is envisioned, wherein the group not only shares its findings and insights but also invites perspectives, critiques, and contributions, ensuring a holistic, multifaceted approach towards understanding and harnessing provenance data effectively.

This section attempts to weave the discussions, explorations, and future directions into a coherent narrative, based on the initial understanding from the provided text files. If there are specific aspects or nuances you'd like to explore or emphasize further, please provide additional guidance or specify areas of deeper interest.

### 4.1.1 Discussions on Definition

This part of the discussion focused on the questions “What really is provenance data?” and “how does it relate to similar concepts?”. This is broken down into a few sub-questions:
1. What can be considered provenance?
2. What are the major use cases to think about for provenance data?
3. Is it related to ontologies? Google has its ontologies, which is a valuable asset for improving product recommendations. Maybe this is also related to the knowledge graphs or knowledge bases?
4. Could we integrate knowledge graphs for visualization recommendation/visual analysis?
5. Maybe this is also related to “Grammars” that formalizes how researchers process and reason about provenance data
   a. Reconciling coarser and finer levels of abstraction for provenance data
   b. Could we formulate grammars to represent interactions?

In the context of human-AI teaming, there are three main provenance components. All of these need to be captured to provide a complete provenance.
1. User’s reasoning/mental model (including granularity)
2. ML’s reasoning/“mental” model (including granularity)
3. Communication between human and ML agents

The availability of provenance has a large impact on possible downstream tasks. The “Imperative” provenance, such as sequences of user interactions, is usually easy to capture, but has limited semantics. “Declarative” provenance, such as user goals, is often more difficult to capture thus less available, but they provide useful insight into the analysis process. The group also observed that there are differences in definitions between the VIS and ML community, and this is to some extent decided by the capabilities of the tool at hand, e.g., limiting supported tasks, available interactions, etc.

### 4.1.2 A possible “Opinionated” Survey Paper

The group had a long discussion about writing an overview or survey paper and decided to start with some possible sections the paper would have and what will be covered in each section. These are detailed below, together with other aspects of the paper the group considered.

#### 4.1.2.1 Preliminaries

These are the assumptions the all the discussions will be based on:
- We consider three main components: humans (users), AI/models, and system (“environment”)
- The provenance captures the history of the “environment”:
  1. Where the human performs the actions the model cares about;
  2. In our case, this is likely an analysis UI/system;
  3. Environment vs Interface vs System: this is a core concept of the paper and the group spend a long time discussing it, which led to the conceptual model shown in Figure 1.

#### 4.1.2.2 Paper search methods

The group then discussed various publication collections that the survey will cover:
1. Review proceedings of VIS, EuroVis, CHI, IUI, FAccT, CG&A, TVCG, TiiS, IVS, CGF,
2. Google scholar search (make sure we record the exact keyword searches we used)
3. LLM search: Abstracts Viewer (`https://pages.graphics.cs.wisc.edu/AbstractsViewer/`) and `http://vitality.mathcs.emory.edu/app`
4. Papers (metadata) collected by Shayan (the conferences and journals covered): AAAI, AISTATS, CGF, CHI, ICML, IUI, KDD, NeurIPS, TiiS, TVCG, VIS (`https://github.com/smonadjemi/mixed-initiative-va-survey/tree/main/paper_data`)
5. Publications from provenance conferences such as Theory and Practice of Provenance

#### 4.1.2.3 What do we consider as provenance?

This closely relates to the scope of the paper. While this paper might not cover every type of provenance, it is useful to have a relatively complete list and then decide what to include. "You might not think it's provenance but it actually is." – Ana. It became clear to some of the group members that their work is related to provenance, but they did not realise that because it was described with a different term. For example, user interaction log is a common type of provenance and ubiquitous in studies evaluating different visual designs and visual analysis systems. However, not everyone realises that this is a form of provenance, and this can be commonplace within the community.

Provenance recording is typically done for a purpose. This has been covered in previous work [20, 27] , but there is a need for further discussion. One fundamental issues is the relationships among the subsets of provenance, as shown in Figure 2. There is the "provenance" in the broadest sense and includes everything that can be captures theoretically. Within this, there is provenance that can be captured practically (the "capturable provenance"), and the ones that no effective recording means exists (such as capturing user thinking). Among the "capturable provenance", some of them are "interpretable", i.e., human can make sense of it. Finally, there is the "relevant provenance", which depends on the application and analysis question, that overlaps all the other types. The group also observed that there is difference between the common interests for academia and industry: while the visualization research community is often more interested in interaction and evaluation, industry tend to care more about data quality, model performance, and governance issues.

There are pros and cons to recording different forms of provenance: Passive/automated log recording does not disrupt the user but can be noisy and lack of meaning, whereas Explicit feedback can be higher quality data for models but disrupts the user's flow. "Passive" Interaction logs include raw system event data and user clicking on typical UI components. However, this also consists of implicit feedback, such as selecting among recommendations, which provides information with richer semantics. This relates to the design idea of "dual purpose interactions", i.e. interactions for performing operations and learning about users. Explicit Feedback, i.e., things you directly ask the user, is less common but it can be very helpful to improve the performance of machine learning models [16]. It is also possible to divide the forms of provenance by its source, i.e., provenance about user and provenance about the system/model. Currently there is no system/tool that combines different types of provenance, especially model+ interaction provenance.

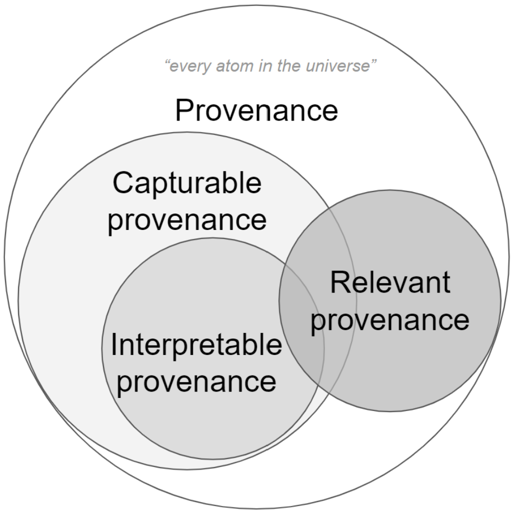



**Figure 2** Different types of provenance and their relationships.

#### 4.1.2.4 Human-AI Interaction

Given the topic of this seminar, human-AI interaction is of particular interest to all the participants. This can be broken down further, as shown in Table 1.

**Table 1** H = high importance, L = low importance.

| Objective | Provenance | ML | User | Data |
|---|---|---|---|---|
| ML Guidance | XAI, Open model | H | L | |
| Guiding ML | Learn form user interaction | L | H | |
| Auditing | Overview, observability | H | H | H |
| Benchmarking | model performance | H | L | H |

#### 4.1.2.5 Representations of Human-AI Interaction

A natural follow-on question is how to represent the provenance of human-AI interaction. Borrowing from the Linguistics community, there can be *semantic* versus *syntactic* relationships between elements. Usually, *grammar* defines syntactic relationships, while *ontology* defines semantic relationships. This has been attempted before for provenance tasks [2], which theorizes a mapping between low-level user interactions and high-level user insights, taking a hierarchical, grammar-like approach. There is also work from outside the visualization community, such as the "Structural summaries for visual provenance analysis" [11] and "Automated Provenance Analytics: A Regular Grammar Based Approach with Applications in Security" [14].

#### 4.1.2.6 Ethical Considerations

There are many ethical considerations related to the collection and use of provenance. One example is *profiling*: the provenance information can be potentially used to profile its source, often a human in this context, for purpose beyond what is originally intended. Related to this is anonymization, i.e., how to remove the personal identifiable information from provenance. An alternative is to seek explicit consent, i.e., user agrees share the information through *data donations*.

## 4.2 Humans

*Jen Rogers (Tufts University – Medford, US) Emily Wall (Emory University – Altanta, US) Mehdi Chakhchoukh (Universite Paris-Saclay, FR) Marie Anastacio (Leiden Universiteit, NL) Rebecca Faust (Virginia Tech – Blacksburg, US) Cagatay Turkay (University of Warwick, GB) Lars Kotthoff (University of Wyoming – Laramie, US ) Steffen Koch (University of Stuttgart, DE) Andreas Kerren (Linköping University, Sweden) Jürgen Bernard (University of Zürich, CH)*



This working group had a total of 10 attendees, who brought diversified perspectives, enriching the discourse. Among the attendees were experts in autoML, and experts in data visualisation. The group focused its efforts on the trade-offs between humans and automation within automated data science. Key areas explored included understanding the AutoML "black box", the role of provenance in supporting diverse user cases, the challenges and affordances of humans and automation, the significance of visualization, and trust-building within AutoML.

The autoML experts showed particular interest in the human interaction aspect at the meta-level, sparking a debate and a categorisation of the main failures that could arise while utilizing autoML. During this categorisation we identified the failures that stem from the automated part of autoML as well as the failures that could stem from the human beings. For instance, is a model's subpar performance an inherent problem to the autoML process or a setup error that comes from human's input?

This led to a deeper discussion about how to unveil the black box surrounding the model and where to include the human in the autoML pipeline. These discussions highlighted provenance as vital to support different user needs, including refinement of user tasks and capturing the rationale for chosen models and their subsequent outcomes. However, further reflection determined that to speculate on provenance, we first needed to define key roles, affordances, challenges of humans, and automation within the data science pipeline as we viewed this as a necessary foundation for further speculation.

Visualization is crucial for helping users understand the process, decisions, and trade-offs within the data science pipeline. Despite its importance in comprehension and trust, the group acknowledged the challenges in visualizing aspects of the process. Additionally, the role of visualization was noted to differ among users, necessitating a nuanced approach tailored to specific stakeholders.

The paradox of human involvement in automated data science remained a consistent theme during our discussions (Figure 3). Do we need humans in the loop at all? What value do they contribute to the data science pipeline? Can we trust a process that we do not understand and are not involved in? Questions such as these provided a catalyst for the group's current efforts in defining tradeoffs between human and automation along the data science pipeline.

Moving forward, the group will continue to develop and refine its perspective on this paradox. This development will include collecting data from the community on their opinions of current and future human-to-machine balances within the data science pipeline through an anonymous online survey. The group will formalize its findings into a written report to contribute to the wider academic community. A potential publication outlet would be the IEEE Computer Graphics and Applications (CG&A) journal.

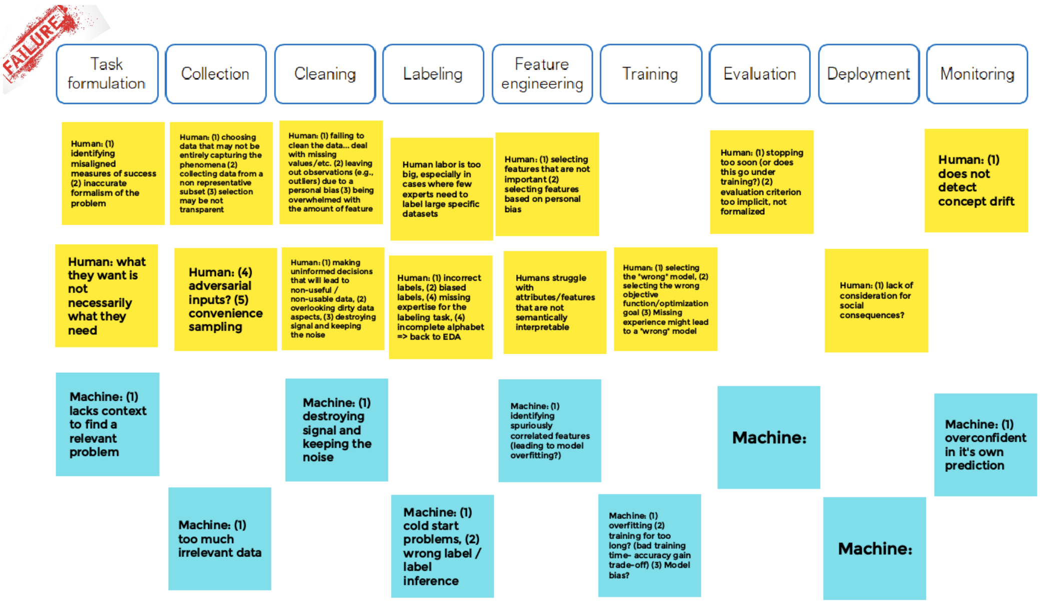



**Figure 3** The Human Paradox of AutoML.

## 4.3 Applications

*Klaus Eckelt (Johannes Kepler University Linz, AT), Sheeba Samuel (Friedrich Schiller University Jena, DE), David Koop (Northern Illinois University – DeKalb, US), Kiran Gadhave (Utah University – Salt Lake City, US), Dominik Moritz (Carnegie Mellon University – Pittsburgh, US)*



Automated Machine Learning (AutoML) and Data Science (AutoDS) pipelines have gained significant attention in recent years. This working group focused on interactive systems to track and visualize the provenance of these pipelines and enable user interaction with automated algorithms as they are running. To gain an overview of the state-of-the-art, we reviewed related work that compares AutoML/DS libraries [5] and applied multiple of them to the same tabular data set. We analyzed the information they provide on the ongoing optimization process, the search space, and their final result (see Table 2). We also compared the resulting models by their complexity and accuracy. All the analyzed libraries provide logs in the console and optionally in a file. Our analysis included Auto-sklearn [7, 6], AutoGluon [4], TPOT [12], FLAML [25], and H2O AutoML [13]. These libraries were used within Jupyter Notebook using Python.[1]

Although some AutoML/DS users have extensive domain knowledge, they may have limited knowledge in the realm of machine learning, and conversely, individuals with a strong machine learning background may lack expertise in the specific domain of the data [3]. As a result, different user groups require different levels of detail and information in visualizing the AutoML/DS process. Related work also differs in terms of detail presented and potential target users. While partial dependence or parallel coordinate plots are easily interpretable,

[1] https://github.com/keckelt/dagstuhl-23372-applications

**Table 2** Overview of the considered AutoML libraries. Downloads were retrieved from PyPI Stats for the last 30 days [8]. The runtime limit was set to 30 minutes. All pipelines were trained with default settings, but maximized CPU utiliznation and logging outputs (in the console (>_) and file (🖹)). ∗ CASH... Combined Algorithm Selection and Hyperparameter Optimization; † PoSH... Portfolio Successive Halving.

| Library | Downloads | Log | Search & Optimization Strategy | Accuracy |
|---|---|---|---|---|
| Auto-sklearn | 25,000 | >_ 🖹 | CASH* with Bayesian Optimization | 95.35 % |
| Auto-sklearn 2.0 | 25,000 | >_ 🖹 | PoSH† and Bayesian Optimization | 93.07 % |
| AutoGluon | 41,000 | >_ 🖹 | Model Portfolio and Random Search | 94.30 % |
| TPOT | 35,000 | >_ 🖹 | CASH with Genetic Programming | 99.92 % |
| FLAML | 189,000 | >_ 🖹 | CASH with Cost-Frugal Optimization | 98.89 % |
| H2O AutoML | 340,000 | >_ | CASH with Random Search | 94.44 % |

tools such as PipelineProfiler [19], ATMSeer [26], or DeepCave [21] allow for a more detailed inspection, but also require technical understanding of the parts of the machine learning pipeline and its optimization.

AutoML/DS approaches and the tools to visualize them currently provide little room for human interaction. The only way to steer the algorithms is by setting parameters before starting an(other) optimization run. But for optimal performance, AutoGluon, for example, advises against human intervention in its documentation[2] and Auto-sklearn 2.0 also removed the human from the loop [6]. Even more steps of the AutoML/DS pipeline will be automated in the future [10].

However, recent research argues for the necessity of reintegrating users into the loop [15]. Given that AutoML/DS requires time to run, it is essential to allow adaptations to make efficient use of this time. Interactive visualization systems can help identify issues early on. For instance, a label left in the training data could cause unusually high performance across all configurations, or poor performance could be due to insufficient data quality. These systems would also enable users to make adjustments to the performance metric, to trade-off between sensitivity and specificity, for example. Providing users with more control when necessary can increase their trust in the AutoML/DS system, often perceived as a 'black box', and could also speed up and improve the process by allowing users to contribute their knowledge more effectively. Meta-learning – i.e., learning from previous experiments – is currently only supported on the machine side of the loop. Auto-sklearn 2.0 [6], for example, looks for similar problems on OpenML [24] to learn from them. However, AutoML/DS systems do not allow users to provide information on similar problems they have encountered. Allowing users to provide their knowledge to the optimization process could guide the search throughout the optimization process.

Figure 4 shows our sketch to visualize AutoML/DS system processes, compare them, and interact with them. Multiple runs can be selected at the top. Their progress and performance metric is displayed on the left side. The large table on the right gives an overview of the configurations that were trained over time. The individual runs are distinguished through different colors (■, ■, ■,). If a configuration fails, we use the negative of the run's color (■, ■, ■). These negatives are less saturated to better differentiate them from successful runs. As there are more configurations than can be displayed on the available vertical screen space, an aggregation method can be selected using the radio buttons on the top right. The best

[2] https://auto.gluon.ai/stable/tutorials/tabular/tabular-essentials.html#maximizing-predictive-performance


**Figure 4** Our sketched visual interface to visualize AutoML/DS systems. An ongoing Auto-sklearn optimization is visualized in green. Additional past runs can be selected at the top for comparison. The table on the right shows the best performing configurations in the specific time segment.

overall configuration per run is additionally highlighted with a colored horizontal bar across the entire line and a trophy symbol next to it. For a detailed inspection, the progress bar, accuracy plot, and table can be vertically expanded using a switch button to show each tested configuration without aggregation. This table can also be used for interaction with the AutoML/DS process. Users should be able to prioritize or block elements to be explored in order to guide the search space. Using an interactive table like LineUp [9] would also allow to filter and rank the configurations, and update performance metrics through combination and weighting of recorded information. However, such an interaction is currently not possible in any of the AutoML/DS systems we reviewed.

We also found that none of the AutoML libraries we tried support MLOps services like MLflow [28] or Weights and Biases [1], which are frequently used to track the training and optimization process of machine learning projects. A custom logging configuration can be passed to Auto-sklearn. All other libraries were only able to log into files instead. We wanted to visualize the AutoML/DS process in real-time while it is running, and thus defined our own logger for Auto-sklearn that also sends all output to Weights and Biases.[3] As this approach heavily relies on log data and parsing string outputs it is limited and error prone. We also noted that these MLOps services do not support the process of tracking AutoML/DS optimizations well, due to the pipeline's many different elements and their parameters.

MLflow or Weights and Biases do not store the recorded information in any standardized or interoperable format, such as PROV-ML [23], which is based on W3C PROV [18]. With *mlflow2prov* [22], data from MLflow and the versioned source code from which it originates can be combined into another provenance format based on W3C PROV.

In addition to the AutoML/DS provenance, a common format to describe the tracked data is necessary. Pipeline elements, their naming, and possible combinations vary between AutoML/DS tools. To ensure that interactive systems are interoperable between AutoML/DS

[3] `https://wandb.ai/dagstuhl-23372/automl`

tools, they require a common standard to communicate and store (intermediate) results [17]. This would allow users to visualize and compare the results of different AutoML/DS systems beyond the performance metrics. We argue that AutoML/DS systems require hooks with which intermediate results are communicated and APIs to steer the ongoing process. We believe the adoption of a standardized provenance format, which can be shared between different data science tools, would facilitate comparisons and allow better monitoring, debugging, interpretation, and explanation of the process.

## References


1 Lukas Biewald. Experiment Tracking with Weights and Biases. `https://www.wandb.com/`, 2020. Accessed: 2023-09-18.

2 Christian Bors, John Wenskovitch, Michelle Dowling, Simon Attfield, Leilani Battle, Alex Endert, Olga Kulyk, and Robert S. Laramee. A provenance task abstraction framework. *IEEE Computer Graphics and Applications*, 39(6):46–60, 2019.

3 Anamaria Crisan, Brittany Fiore-Gartland, and Melanie Tory. Passing the Data Baton: A Retrospective Analysis on Data Science Work and Workers. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1860–1870, 2021.

4 Nick Erickson, Jonas Mueller, Alexander Shirkov, Hang Zhang, Pedro Larroy, Mu Li, and Alexander Smola. AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data. *arXiv preprint arXiv:2003.06505*, 2020.

5 Luís Ferreira, André Pilastri, Carlos Manuel Martins, Pedro Miguel Pires, and Paulo Cortez. A Comparison of AutoML Tools for Machine Learning, Deep Learning and XGBoost. In *2021 International Joint Conference on Neural Networks*, pages 1–8, July 2021.

6 Matthias Feurer, Katharina Eggensperger, Stefan Falkner, Marius Lindauer, and Frank Hutter. Auto-sklearn 2.0: hands-free AutoML via meta-learning. *The Journal of Machine Learning Research*, 23(1):261:11936–261:11996, January 2022.

7 Matthias Feurer, Aaron Klein, Katharina Eggensperger, Jost Springenberg, Manuel Blum, and Frank Hutter. Efficient and Robust Automated Machine Learning. In *Advances in Neural Information Processing Systems 28 (2015)*, pages 2962–2970, 2015.

8 Christopher Flynn. PyPI Download Stats. `https://pypistats.org/`, 2023. Accessed: 2023-09-18.

9 Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. LineUp: Visual Analysis of Multi-Attribute Rankings. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2277–2286, 2013.

10 Shubhra Kanti Karmaker (“Santu”), Md. Mahadi Hassan, Micah J. Smith, Lei Xu, Chengxiang Zhai, and Kalyan Veeramachaneni. AutoML to Date and Beyond: Challenges and Opportunities. *ACM Computing Surveys*, 54(8):175:1–175:36, 2021.

11 Houssem Ben Lahmar and Melanie Herschel. Structural summaries for visual provenance analysis. In *Proceedings of the 11th USENIX Conference on Theory and Practice of Provenance*, TAPP’19, page 2, USA, 2019. USENIX Association.

12 Trang T Le, Weixuan Fu, and Jason H Moore. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics*, 36(1):250–256, 2020.

13 Erin LeDell and Sebastien Poirier. H2O AutoML: Scalable Automatic Machine Learning. *7th ICML Workshop on Automated Machine Learning*, July 2020.

14 Mark Lemay, Wajih Ul Hassan, Thomas Moyer, Nabil Schear, and Warren Smith. Automated provenance analytics: A regular grammar based approach with applications in security. In *Proceedings of the 9th USENIX Conference on Theory and Practice of Provenance*, TaPP’17, page 12, USA, 2017. USENIX Association.

**15** Marius Lindauer and Alexander Tornede. Rethinking AutoML: Advancing from a Machine-Centered to Human-Centered Paradigm. `https://www.automl.org/rethinking-automl-advancing-from-a-machine-\centered-to-human-centered-paradigm/`, 2022. Accessed: 2023-09-18.

**16** Yannick Metz, David Lindner, Raphaël Baur, Daniel Keim, and Mennatallah El-Assady. RLHF-Blender: A Configurable Interactive Interface for Learning from Diverse Human Feedback, August 2023. arXiv:2308.04332 [cs].

**17** Mitar Milutinovic. *Towards Automatic Machine Learning Pipeline Design*. PhD thesis, EECS Department, University of California, Berkeley, Aug 2019.

**18** Paolo Missier, Khalid Belhajjame, and James Cheney. The W3C PROV family of specifications for modelling provenance metadata. In *Proceedings of the 16th International Conference on Extending Database Technology*, pages 773–776, New York, USA, 2013. ACM.

**19** Jorge Piazentin Ono, Sonia Castelo, Roque Lopez, Enrico Bertini, Juliana Freire, and Claudio Silva. PipelineProfiler: A Visual Analytics Tool for the Exploration of AutoML Pipelines. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):390–400, February 2021.

**20** E.D. Ragan, A. Endert, J. Sanyal, and J. Chen. Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes. *IEEE Transactions on Visualization and Computer Graphics (VAST '15)*, 22(1):31–40, 2016.

**21** René Sass, Eddie Bergman, André Biedenkapp, Frank Hutter, and Marius Lindauer. Deep-CAVE: An Interactive Analysis Tool for Automated Machine Learning. In *ICML Workshop on Adaptive Experimental Design and Active Learning in the Real World*, June 2022.

**22** Marius Schlegel and Kai-Uwe Sattler. Extracting Provenance of Machine Learning Experiment Pipeline Artifacts. In Alberto Abelló, Panos Vassiliadis, Oscar Romero, and Robert Wrembel, editors, *Advances in Databases and Information Systems*, Lecture Notes in Computer Science, pages 238–251. Springer Nature Switzerland, 2023.

**23** Renan Souza, Leonardo Azevedo, Vítor Lourenço, Elton Soares, Raphael Thiago, Rafael Brandão, Daniel Civitarese, Emilio Brazil, Marcio Moreno, Patrick Valduriez, Marta Mattoso, Renato Cerqueira, and Marco A.S. Netto. Provenance Data in the Machine Learning Lifecycle in Computational Science and Engineering. In *2019 IEEE/ACM Workflows in Support of Large-Scale Science*, pages 1–10, November 2019.

**24** Joaquin Vanschoren, Jan N. van Rijn, Bernd Bischl, and Luis Torgo. OpenML: networked science in machine learning. *SIGKDD Explorations*, 15(2):49–60, 2013.

**25** Chi Wang, Qingyun Wu, Markus Weimer, and Erkang Zhu. FLAML: A Fast and Lightweight AutoML Library. *Proceedings of Machine Learning and Systems*, 3:434–447, 2021.

**26** Qianwen Wang, Yao Ming, Zhihua Jin, Qiaomu Shen, Dongyu Liu, Micah J. Smith, Kalyan Veeramachaneni, and Huamin Qu. ATMSeer: Increasing Transparency and Controllability in Automated Machine Learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, New York, USA, 2019. ACM.

**27** Kai Xu, Alvitta Ottley, Conny Walchshofer, Marc Streit, Remco Chang, and John Wenskovitch. Survey on the Analysis of User Interactions and Visualization Provenance. *Computer Graphics Forum*, 39(3):757–783, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14035.

**28** Matei Zaharia, Andrew Chen, Aaron Davidson, Ali Ghodsi, Sue Ann Hong, Andy Konwinski, Siddharth Murching, Tomas Nykodym, Paul Ogilvie, Mani Parkhe, and others. Accelerating the Machine Learning Lifecycle with MLflow. *IEEE Data Engineering Bulletin*, 41(4):39–45, 2018.

## Participants


- Marie Anastacio
RWTH Aachen, DE
- Leilani Battle
University of Washington – Seattle, US
- Jürgen Bernard
Universität Zürich, CH
- Nadia Boukhelifa
INRAE – Palaiseau, FR
- Camelia D. Brumar
Tufts University – Medford, US
- Mehdi Chakhchoukh
Université de Paris-Saclay – Gif-sur-Yvette, FR & INRAE – Paris, FR
- Anamaria Crisan
Tableau Software – Seattle, US
- Klaus Eckelt
Johannes Kepler Universität Linz, AT
- Mennatallah El-Assady
ETH Zürich, CH
- Alex Endert
Georgia Institute of Technology – Atlanta, US
- Rebecca Faust
Virginia Polytechnic Institute – Blacksburg, US
- Kiran Gadhave
University of Utah – Salt Lake City, US
- Andreas Kerren
Linköping University, SE
- Steffen Koch
Universität Stuttgart, DE
- David Koop
Northern Illinois University – DeKalb, US
- Lars Kotthoff
University of Wyoming – Laramie, US
- Alexander Lex
University of Utah – Salt Lake City, US
- Dominik Moritz
Carnegie Mellon University – Pittsburgh, US
- Alvitta Ottley
Washington University – St. Louis, US
- Jen Rogers
Tufts University – Medford, US
- Sheeba Samuel
Friedrich-Schiller-Universität Jena, DE
- Marc Streit
Johannes Kepler Universität Linz, AT
- Tanja Tornede
Leibniz Universität Hannover, DE
- Cagatay Turkay
University of Warwick – Coventry, GB
- Emily Wall
Emory University – Atlanta, US
- Kai Xu
University of Nottingham, GB


Report from Dagstuhl Seminar 23381

# Visualization and the Humanities: Towards a Shared Research Agenda

**Johanna Drucker*[1], Mennatallah El-Assady*[2], Uta Hinrichs*[3], Florian Windhager*[4], and Derya Akbaba†[5]**

1 **University of California at Los Angeles, US.** drucker@gseis.ucla.edu
2 **ETH Zürich, CH.** melassady@ai.ethz.ch
3 **University of Edinburgh, GB.** uhinrich@ed.ac.uk
4 **Donau-Universität Krems, AT.** florian.windhager@donau-uni.ac.at
5 **Linköping University, SE**

**Abstract**

This report documents the program and the outcomes of Dagstuhl Seminar 23381 *"Visualization and the Humanities: Towards a Shared Research Agenda"*. The seminar was motivated by the fact that visualization has become a vital element in (digital) humanist research practices recently, while the value and impact of research at the intersection of visualization and the humanities is still widely debated and frequently contested from both sides. Visualization scholars critique the service-oriented focus on visualization as a tool to facilitate humanist research, which hampers the discovery of complementary and mutually enriching research perspectives for all fields involved. At the same time, humanists warn of visualizations' roots in the quantitative sciences which introduce non-trivial shifts in the topology of knowledge-power, creating epistemic, political, ethical, pedagogical, and cultural tensions. Building on advances in this young and highly interdisciplinary research area, the seminar discussed how to leverage synergies and how to build productively on tensions between methodologies at the intersection of visualization and (digital) humanities fields that span a vast spectrum of research philosophies and methods. The seminar thus brought together researchers and practitioners from the fields of visualization, computer science, the humanities, and design to reflect on existing research methods within visualization and the humanities, to identify tensions and synergies between the different fields, and to develop concrete avenues that address and leverage these.

**Seminar** September 17–22, 2023 – https://www.dagstuhl.de/23381

**2012 ACM Subject Classification** Human-centered computing → Visualization design and evaluation methods; Applied computing → Arts and humanities; Human-centered computing → Visualization theory, concepts and paradigms; Human-centered computing → Human computer interaction (HCI)

**Keywords and phrases** Digital humanities, arts, humanities, methodology, research program, visualization

**Digital Object Identifier** 10.4230/DagRep.13.9.137

* Editor / Organizer
† Editorial Assistant / Collector



Visualization and the Humanities: Towards a Shared Research Agenda, *Dagstuhl Reports*, Vol. 13, Issue 9, pp. 137–165
Editors: Johanna Drucker, Mennatallah El-Assady, Uta Hinrichs, and Florian Windhager
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

# 1 Executive Summary

*Johanna Drucker (University of California at Los Angeles, US)*
*Mennatallah El-Assady (ETH Zürich, CH)*
*Uta Hinrichs (University of Edinburgh, GB)*
*Florian Windhager (Donau-Universität Krems, AT)*



Large-scale digitization initiatives of cultural materials have fueled an interest in both computer science and humanist research fields to develop advanced computational methods specifically tailored to humanities research. In this context, recent years have seen a particular excitement around visualization (VIS) as a method to facilitate humanist research practices, visible in a wave of surveys and reflections from both humanist and visualization perspectives (Benito-Santos, 2020; Bradley, 2018; Drucker, 2020; Hinrichs, 2019; Jänicke, 2017; Windhager, 2018).

From a digital humanities' (DH) perspective, visualization appears as a fascinating option to make sense of cultural collections and subject matters at a large scale. Visualization has the potential to complement traditional humanities (TH) research approaches that typically focus on the in-depth interpretation of selected cultural materials. From the perspective of visualization researchers, the humanities provide a unique application domain for designing and studying the impact of visualization tools and related sense-making processes. Maybe more importantly, however, the humanities stimulate methodological and conceptual innovation in the visualization field, by formulating novel demands on visualization while criticizing standard visualization approaches that have their roots in the sciences. However, conceptualizing and conducting impactful research that mutually benefits and fuels all fields involved in this highly interdisciplinary area is an enormous challenge. Visualization scholars depend on humanities experts to adapt and tailor technologies for a complex and critical application field, but they also suffer from a dominant unidirectional focus on visualization 'as a service', which hampers the development of novel, complementary and mutually enriching research perspectives (Hinrichs, 2019). At the same time, many humanists warn of visualizations' roots in the quantitative sciences which disrupts traditional humanist methods of knowledge generation and discourse and introduces non-trivial shifts in existing topologies of knowledge-power, creating epistemic, political, ethical, pedagogical, and cultural tensions (Allington, 2016; D'Ignazio, 2020; Marche, 2012); scholars have warned of intellectual *Trojan horses* (Drucker, 2011), the *curse of counting* (Da, 2019), and *unhappy neighborhoods* (Correll, 2019).

Visualization and the humanities, indeed, seem to subscribe to fundamentally different research philosophies, approaches and methods. There is also a lack of explication and understanding of what characterizes methodologies across the fields and how to manage the complex mixture of methodological similarities and disparities. This not only challenges research collaborations at the intersection of visualization and the humanities and complicates the identification of core foundations and methods to teach, but also hampers the formulation of research outcomes and a shared research agenda to drive forward future interdisciplinary initiatives in this area. In order to unleash and direct synergies inherent in transdisciplinary research and teaching across the humanities and visualization we need to share, converse and reflect on each others' research approaches, methods and processes and how these impact and shape research at the intersection of the humanities and visualization. Along these lines, this Dagstuhl Seminar aimed at addressing three main questions and related challenges:

- **DH⇒*VIS: How do humanistic approaches impact visualization research and practice?***
- **VIS⇒*DH: How does visualization impact knowledge production in the humanities?***
- **DH⇔*VIS: Leveraging synergies between the fields, what could a shared research agenda look like?***

## References


1 Allington, D., Brouillette, S., and Golumbia, D., 2016. Neoliberal tools (and archives): A political history of digital humanities. LA Review of Books.
2 Benito-Santos, A., and Sánchez, R. T., 2020. A data-driven introduction to authors, readings, and techniques in visualization for the digital humanities. IEEE Computer Graphics and Applications 40(3), pp.45–57.
3 Bradley, A. J., El-Assady, M., Coles, K., Alexander, E., C. M., Collins, C., Jänicke, S., and Wrisley, D. J., 2018. Visualization and the digital humanities: Moving towards stronger collaborations. IEEE Computer Graphics and Applications 38(6), pp.26–38.
4 Correll, M., 2019. Counting, collaborating, and coexisting: Visualization and the digital humanities. online:, 2019.
5 Da, N. Z., 2019. The Computational Case against Computational Literary Studies. Critical Inquiry 45(3), pp. 601–639. Publisher: The University of Chicago Press.
6 Drucker, J., 2011. Humanities approaches to graphical display. Digital Humanities Quarterly 5(1).
7 Drucker, J., 2020. Visualization and Interpretation: Humanistic Approaches to Display. MIT Press.
8 D'Ignazio, C., and Klein, L. F., 2020. Data Feminism. MIT Press.
9 Hinrichs, U., Forlini, S., and Moynihan, B., 2019. In defense of sandcastles: Research thinking through visualization in digital humanities. Digital Scholarship in the Humanities 34, Supplement 1, i80–i99.
10 Jänicke, S., Franzini, G., Cheema, M. F., and Scheuermann, G., 2017. Visual text analysis in digital humanities. Computer Graphics Forum 36(6), pp. 226 – -250.
11 Marche, S., 2012. Literature is not data. against digital humanities. LA Review of Books.
12 Windhager, F., Federico, P., Schreder, G., Glinka, K., D. M., Miksch, S., and Mayr, E., 2018. Visualization of cultural heritage collection data: State of the art and future challenges. Transactions on Visualization and Computer Graphics 25(6), pp. 2311—2330.

## 2 Table of Contents

# 3 From Motivation to Realization

Directly building upon and advancing related ongoing discussions in visualization and the humanities, this Dagstuhl Seminar brought together researchers and practitioners from the fields of visualization, computer science, the humanities, and design with the goal of outlining and concretizing methodological synergies to fuel novel research that has an impact on all research fields involved. In particular, the aims of the seminar were to:

- Reflect on existing research methods within visualization and the humanities,
- Identify tensions and synergies between the different fields, and
- Develop concrete avenues that address and leverage these.

By the end of the seminar, these goals had been achieved by supporting active and projected ongoing work in these areas:

1. Community building within the interdisciplinary group in order to sustain future research;
2. Formation of subgroups to support exchange of information and sustain collaboration;
3. Creating specific outputs for publication and resource building;
4. Targeting projects and activities for short and longer-term research activity.

### Seminar Structure & Activities

The seminar was structured into 4 parts and related activities: a *pre-seminar reflection*, an *icebreaker* activity, a *topic brainstorming* activity, *input talks* along the three highlighted questions above, and corresponding *breakout sessions* in working groups and cross-cutting *plenum discussions*.

**Pre-seminar Reflection.** In order to get participants thinking about the seminar's topic and challenges, we invited them a few weeks prior to identify and characterize research methods/theories/approaches that they felt were relevant to the fields of visualization and the humanities. We facilitated this reflection activity in the form of a brief survey that invited participants to explore the following prompts:

- Choose and name one method/theory/approach that you find relevant for knowledge production in the area of your expertise.
- Do you think this method/theory/approach belongs to a certain or multiple field(s) or area(s)? If so, please specify.
- What are the key characteristics of this method/theory/approach?
- Why do you deem the method/theory/approach as interesting? How have you used it? Why does it have promise?
- How well is knowledge – developed using this approach – received? How reliably is it judged?
- How are successes and failures of applying this method/theory/approach characterized?
- What is the relation of this method/theory/approach to the humanities (if any)?
- What is the relation of this method/theory/approach to visualization (if any)?
- How does this method/theory/approach create or leverage synergies between visualization and the humanities (if at all)?
- Please add any references below that illustrate the method/theory/approach and its application.

The idea of this survey was to help establish a common understanding of relevant concepts – and to discuss inter-method relations on that basis. Participants submitted over 20 approaches that ranged from quantitative and qualitative methods for data analysis, theories, as well as practical design methods. These formed a valuable basis for break-out group discussions (see, for example, the summary of the *Theories* & *Methods* working group).

**Icebreaker Activity.** One of the challenges we anticipated for the seminar was to connect researchers and practitioners from quite diverse fields and to create an environment that would enable participants to engage in in-depth, sometimes controversial, yet productive discussions. In order to address this, one of the first seminar activities consisted of an icebreaker activity which, initially developed by Tatiana Losev[1], invited participants to create a sketch using pencils and paper reflecting on the following prompt:

**How do I see myself in relation to the seminar topic: "Visualization and the Humanities: Towards a Shared Research Agenda"**

Rather than using pre-fabricated slides to introduce themselves, participants introduced themselves through their sketches (see figure below). This activity already revealed the many different perspectives participants brought to the seminar topics as well as a remarkable diversity of possible representation approaches – some playful, some practical, some more abstract.

**Topic Brainstorming.** The icebreaker activity was followed by a topic brainstorming that invited participants to identify topics, questions, challenges, gaps, barriers and activities that they would like to see being discussed during the seminar. This activity followed a traditional "collection & clustering" approach where topics were first gathered in the form of sticky notes and then collectively clustered by theme, again, roughly following the three perspectives VIS⇒DH, DH⇒VIS, and synergies between VISDH⇔DH. These topic clusters, along with the input talks, formed the basis for discussions in break-out groups (for an overview of selected topics discussed, see the summary of working groups).

**Input Talks.** Throughout the seminar we had three sessions of semi-formal input talks that were roughly structured by perspective (VIS⇒DH, DH⇒VIS, and synergies between DH⇔VIS) and that focused on topics central to the shared research agenda: visualization, ethics, methods in the humanities, pedagogical concerns, and intersections of epistemic, political and cultural tensions in visualization as it intersects with the humanities. These input sessions provoked lively and constructive discussion that, in turn, fed into the break-out sessions. An overview of the input talks can be found in the **Overview of Talks** section of this report.

[1] https://www.tatianalosev.com/

**Break-out Sessions.** The themes for these sessions were generated in the topic brainstorming mentioned above through use of clustered post-its and synthesis of questions and topics they contained. Break-out groups ranged in size and were flexible, so participants could move among them. An overview of topics discussed in break-out groups can be found in the **Working Groups** section of this report.

**Plenum Discussions.** These were essential conversations that involved all seminar participants and allowed consensus and agenda-setting to emerge from conversations.

Energies never flagged in the course of the week, and sub-group conversations and research interests continued to form through the meal-time and coffee-break exchanges, as well as in the after-hours lounge. One evening was used to take up a prompt from one of the break-out groups and do a drawing exercise with volunteer participants (see *Vis in the Humanities & Encodings (output session)*. Two-thirds of the seminar attendees engaged in the activity, with positive results that may become the foundation of a publication analyzing the potential for innovation in the outcomes.

### Methods and Platforms

Several tools and platforms were used before, during, and after the seminar to provide shared resources, real-time responses and dialogue, as well as record-keeping. These included:

- **Miro Board** for sharing materials: This was perhaps the least used of these platforms.
- **Joint Note-taking** in Google Drive: This was highly used and useful in real time and for creating a record of presentations, and equally important, responses and discussions.
- **Discord Channel:** Supports lively, ongoing real-time exchanges, as a way to keep sub-groups in touch and the Seminar's conversation going.
- **Google Drive:** Proved to be invaluable for collecting inputs, reports, abstracts, and other materials useful as a record of the seminar and also as a resource for future work.

### Reflections on the Planning Process

The planning process unfolded over an almost three-year period, with several iterations resulting in a successful proposal. The five coordinators (including Jason Dykes who was unable to attend the seminar) met regularly on Zoom to define the motivation, methods, structure, schedule, and potential participants. The combination of expertise and variety of professional networks was reflected in the highly interdisciplinary composition of the seminar participants. No recommendations for changes in this process seem necessary, though an account of the roles and responsibilities, even tasks that need attention at different points in the planning timeline, might assist future organizers in being sure that there is an even distribution of the workload. Our group functioned extremely well in this regard.

### Outline of Outcomes

The outcomes of the seminar included two major areas of activity: a) *community building* and b) developing *topic-specific output goals.*

**Community Building.** The seminar identified a large number of topics that require further discussion and engagement and exchanges to consolidate and develop the VIS+DH field of practice. While initiatives such as the VIS4DH workshop exist, the seminar confirmed a clear need for further community building to sustain and deepen a whole range of interdisciplinary exchanges at the intersection of Visualization and Digital Humanities. Related Seminar discussions resulted in:

- The commitment to create a website for sharing resources, projects, calls for papers and emerging research opportunities;
- An email list for posting events and activities;
- A writing group to review and facilitate publications at the intersection of visualization and the humanities;
- Various plans for future meetings and coordination of individual working groups (see **Working Groups** section in this report) – coordinated mainly through the Seminar's Discord channel.

**Topic specific outputs.** Each working group defined their own potential collcetion of outputs (see **Working Groups** section), including but not limited to:

- Plans for an anthology of short-form reports on various topics (see topics discussed by working groups);
- Ideas for papers and longer-form research on ethics, methods, practices, and assessment criteria;
- Possibilities for research collaboration from within each group;
- Interest in creating pedagogical resources.

### Suggestions for Future and Follow-up Activities

The seminar generated tremendous intellectual and professional excitement about collaboration on research, pedagogical resource development, and sustained activities at the intersection of visualization and the humanities. A workshop that would support rapid prototyping of proposed projects to see what would be required from the visualization, computer science, and design communities and from the humanities' community might expose more about how such work would support genuinely rewarding intellectual activity with potential for applications and implementation with broader use. In particular, the idea of interactive visualizations that promote and sustain interpretative practices, predictive models, transformative tools, and other innovations might be explored with the goal of imagining their use in business, scientific, educational, social science, and cultural/institutional domains.

# 4 Overview of Talks

## 4.1 What are Methods in the Humanities?

*Johanna Drucker (University of California at Los Angeles, US)*



Only one method exists in the humanities: interpretative reading. The method has many modalities (approaches) and techniques (applications) but within the historical tradition from which the humanities derive, the act of reading (mainly) texts is the single approach to knowledge and understanding. This can be tracked in Western culture to the reading of sacred texts in the religious communities of the three Abrahamic faiths: Judaism, Christianity, and Islam. This practice is also central to Eastern traditions in Buddhism, Hinduism, and Confucianism. While the study of images, dance, spatial rituals, and other activities are persistent in humanistic practices, the development of disciplines that considered these as objects of study comes much later than textual study. As a field, art history appears in the late 18th-century. The social sciences such as sociology, anthropology, archaeology are largely the outcome of 19th century syntheses of empirical methods from natural sciences and humanistic topics of research. By analogy, these disciplines perform "readings" of images, practices, and social phenomena that are not constituted solely by texts. The various inflections that nuance the method of interpretative reading – critical theory, deconstruction, queer theory, feminism, decolonizing approaches – each bring a set of distinct and significant perspectives and insights into the ideological frame, calling attention to often unacknowledged assumptions and biases of traditional interpretation – but these are also interpretative reading practices.

## 4.2 Scraps

*Stefania Forlini (University of Calgary, CA) and Bridget Moynihan (Library and Archives Canada – Ottawa, CA)*



With this talk, we enter ongoing traditions of salvaging, repurposing, and re-valuing the remnants of the past. Drawing predominantly from feminist, queer, and archival critical theories, as well as artistic and/or vernacular practices, we offer what we call a "scavenger method". We focus on scraps that preserve (albeit imperfectly, in different degrees of ruin) material traces of other times and "users". These scraps offer opportunities to read history "against the grain" of dominant narratives and work as a provocation for visualization and humanities researchers to experiment with practices of re/contextualization of historically specific data.

## 4.3 Simple VIS Makes LIT Complex (As they should)

*Christophe Schuwey (Université de Bretagne Sud, FR)*



Scholars in the Humanities often think of data visualization as graphs, maps and trees. In this talk, I argue for the essential role of other, less expected types of visualization, and how humanities hypothesis can extend and diversify the notion of data and visualization, both for the fields of humanities and data visualization. Indeed, for scholars in Early Modern literature, one example of visualization revolutions lies in the ability to access and read Early Modern texts in their original layout, thanks to initiative such as Google Books or Gallica. Considering type size, layout and text formatting as data reintegrate issues of materiality directly into textual analysis, challenging traditional approaches and methods in literary studies. I used the example of La Bruyòre's Caractères to show how several interfaces designed with a fine-grained understanding of the work in mind enable the work to be visualized differently, profoundly transforming the understanding and perception of a literary masterwork. Such interfaces challenge the preconceived idea of a marmoreal masterpiece, from restoring its inherent dynamics (multiple editions with multiple additions and modifications over a short period), to enabling new approaches to its various translations across Europe.

## 4.4 Computational Iconographies

*Fabian Offert (University of California – Santa Barbara, US)*


**Joint work of** Leonardo Impett, Fabian Offert
**Main reference** Leonardo Impett, Fabian Offert: "There Is a Digital Art History", CoRR, Vol. abs/2308.07464, 2023.
**URL** https://doi.org//10.48550/ARXIV.2308.07464

While more traditional types of neural networks have long been part of digital art history, the epistemic implications and methodological affordances of contemporary large-scale vision models have not yet been systematically analyzed. We focus our analysis on two main aspects that, together, seem to suggest a coming paradigm shift towards a "digital" art history in Johanna Drucker's sense. On the one hand, the visual-cultural repertoire newly encoded in large-scale vision models has an outsized effect on digital art history. The inclusion of significant numbers of non-photographic images allows for the extraction and automation of different forms of visual logics, from which computational iconographies of almost arbitrary complexity emerge. Large-scale vision models have "seen" large parts of the Western visual canon mediated by Net visual culture, and they continuously solidify and concretize this canon through their already widespread application in all aspects of digital life. On the other hand, based on a technical case study of utilizing a contemporary large-scale visual model to investigate basic questions from the field of urbanism, we suggest that such systems require a new critical methodology that takes into account the epistemic entanglement of a model and its applications. This new methodology reads its corpora through a neural model's training data, and vice versa: the visual ideologies of research datasets and training datasets become entangled.

## 4.5 Data Visualization in the Humanities – Challenges and Opportunities

*Steffen Koch (Universität Stuttgart, DE)*


**Joint work of** Steffen Koch, Max Franke, Dorothea Weltecke, Ralph Barczok

My presentation discusses conditions under which research scenarios in the humanities can benefit from abstract data visualization. Using the Damast prototype as an example, the talk demonstrates how the combination of abstract data visualization, the recording of analysis provenance, and long-term data preservation support the reproducibility of interactive visual analyses. However, even when providing a certain level of reproducibility, it is necessary to provide sufficient context when presenting visually abstracted research data. An orthogonal topic are cultural differences between research disciplines in collaborative projects. Publication traditions and the reception of joint research in the individual disciplines can negatively impact early-stage research careers. Another issue are the different requirements regarding the sustainability of research outcomes. If collaborative research should become a success story, these aspects need to be discussed openly.

## 4.6 Visualization & Uncertainty: A Love/Hate Relationship

*Michael Correll (Northeastern University – Portland, US)*



The reification of data is the bedrock of visualization. Uncertainty adds roadblocks to this process: how do we chart the unknown or the unknowable? Overcoming this roadblock involves the reduction of uncertainty down to probabilities, a "real abstraction" that fails to capture how people experience or reason about uncertainty, but nevertheless has normative force: we often try to make people act as "statistical golems", and punish/degrade them when they don't. In this talk I explore the frictions introduced by the probability-based visualization of uncertainty, and the visualization work that highlights these frictions.

## 4.7 Epistemology of History Research as Seen by a Visualization Researcher

*Jean-Daniel Fekete (INRIA Saclay – Orsay, FR)*



Visualization researchers are very familiar with the epistemology of natural sciences and follow its model for exploring data. The social sciences do not follow this epistemology, leading to misunderstandings between the two communities when discussing how visualization tools should support research. There is a need to align the expectations and assumptions between the two communities to converge to usable tools and more constructive criticisms. In particular, the social sciences deal with several levels of beliefs for their research questions; some hypotheses are plausible while others are competing and can remain with these statuses for a long time (for generations). Additionally, the nature of the "data" is not as clear-cut as in the natural sciences. Therefore, this talk is an invitation to visualization researchers and practitioners to better support the subtle levels of truthfulness required by the social sciences as well as mechanisms to intervene in the data to provide contextual information (in a transparent manner).

## 4.8 Best Practices Considered Harmful (some of the time)

*Sheelagh Carpendale (Simon Fraser University – Vancouver, CA)*



I will start this talk by discussing the differences and nuances among the concepts of Best Practices, Guidelines, Lessons Learned, and Heuristics, pointing out frequently ignored pitfalls and potentially looming pain points. I will then discuss how the pendulum between advantages and disadvantages can be mitigated by individual attitudes and actions.

## 4.9 Unvisualizing Texts: Hermeneutics of Visualization in Textual Scholarship

*Joris van Zundert (Huygens Institute – Amsterdam, NL)*



In this contribution, I discuss the suboptimal visualizations that textual scholarship applies to transmedialize (historical) text sources in a digital environment. Current digital scholarly editions are text (or glyph) essentialist and ignore for the most part all material information of the documents that they are contained in, ignoring thus also all aspects of non-glyph communication besides sometimes relational information expressed by e.g. marginals, footnotes, etc. This current TEI-XML "good practices" based approach has hermeneutical and pernicious epistemological ramifications as digital scholarly editions are repeating indefinitely a book metaphor, but in a bland abstracted, reductive, and non-creative way. Hermeneutically, this does a disservice to the texts and documents they are trying to re-represent. Epistemologically, they are relatively "poor" and do not inspire exploration of any new engagement or knowledge inference from the texts themselves.

## 4.10 Some synergies: Working together at intersections of Visualization & (Digital) Humanities

*Jeremy Douglass (University of California – Santa Barbara, US)*



Looking beyond the model of formal research collaborations between disciplined visualization researchers & humanities scholars, this input reflects on expanded ideas of synergy (sun+ergos) as "working together." In role-based thinking, working together may include a humanities scholar + visualization researcher, or artist... or a hum scholar who is ALSO a vis researcher / artist (& vice versa). In work-based thinking, beyond the "two halves" work model ("I do vis, you do humanities") is instead the "speak the same language" model: e.g. this specific intersection of the humviz is a domain, so we develop a shared domain-specific language (in a technical or general sense). In each community you afford a different "you" for purposes of synergy – my past collaborative "mes" include artist, domain expert, developer et cetera. Some of the most productive thinking about "working together" is grounded in personal experiences, situating ourselves and reflecting on projects and roles, moving between e.g. the humanities, information visualization, computer science, & fine arts. We collaborators are usually more complex than the simplified role they overtly play, and have more to offer a collaboration. I explore this idea through role-based reflection on four of my visualization+humanities synergies at different career stages: 1) an artist researching historical visualization forms as traditions within community (as PhD researcher, Software Studies Initiative), 2) a Critical Code Studies hermeneuticist using artist-provocateur exploratory visualization (as co-author on Reading Project: A Collaborative Approach to Digital Literary Criticism), 3) the "infrastructure guy" supporting humanities teams with virtual machine clouds while they use visualization to read (grant Co-PI on the WhatEvery1Says Project), and 4) teaching viz to humanities students (as faculty lab director / project lead) in two projects: Panelcode, a minimal markup language for visualizing comics compositions as abstract layouts, and in The Transverse Reading Project, a visual atlas of branching narratives. When "working together" in hum+viz, reflect on your experiences and situate yourself. When seeking synergies, which "you" is the collaborator?

## 4.11 Electronic Health Records that support clinical reasoning

*David Pao (Royal College of Art – London, GB)*


**URL** https://users.sussex.ac.uk/ peterch/papers/ComplexProbCh05.pdf

Unlike the paper medical record, which is created by clinicians for their own use, the Electronic Health Record (EHR) is predominantly designed by non-clinicians. The EHR interface is the cornerstone tool through which clinicians review, interpret, create and curate patient data. EHR interfaces are consistently criticized for their poor usability, which results in clinician burnout and sub-optimal, clinically misaligned data being propagated throughout a health system–infinitely and without degradation.

This research aims to realign EHR interface design with the central tenet of user-centredness by better understanding the clinician user. It seeks to achieve this by understanding how the EHR interface can specifically support a clinician's practice–defined here for the first time as clinical usability (CU).

The hypothesis is that, using a provocative prototype designed by a single clinician as a visual starting point, a Participatory Action Research (PAR) approach can effectively capture, codify and communicate a clinician community's CU knowledge to inform the trans-disciplinary field of EHR interface design. This leads to the central research question, "How can a clinician community contribute CU knowledge to the design of their EHR interface?"

This research contributes tangible, real-world and communicable new knowledge. First, CU knowledge within a set of novel CU-specific heuristics, which bridge the longstanding gaps between the designer and clinician user, and design and evaluation, that have come to characterize this design field. Secondly, CU knowledge embodied within the Sexually Transmitted Infection Query Interface (STIQI) development prototype and its Representational Epistemic (REEP) design blueprint, not previously seen in either commercial or research settings.

## 4.12 Towards radically thick new tools: Visualizing ambiguity

*Mathieu Jacomy (Aalborg University Copenhagen, DK)*



Data visualization, as a practice and as an academic field, has dedicated more time to visualizing uncertainty than ambiguity. Those are not the same thing: although knowledge reduces uncertainty, it amplifies ambiguity. Ambiguity arises as a feature of the things, beings and phenomena we describe and study.

In this short talk, I draw on a minimal case: community detection in networks. Although we can say that communities exist in the network I use for an example, their demarcations are ambiguous. I show five different ways to visualize that ambiguity, and I propose a scale to help people assess the status of ambiguity in their own visualizations, or that of others.

1. Ambiguity as noise: it is not visualized
2. Ambiguity as an accident: visualized as an anecdote
3. Ambiguity as context: it is visualized, but as an afterthought
4. Ambiguity as a problem: it is fully visualized, but gets in the way of understanding the rest
5. Ambiguity as a feature: it is the purpose of the visualization.

From there, I sketch a possible program at the intersection of the digital humanities and visualization: building radically thick new tools. It may draw on two influences. The "thick description" formalized by anthropologist Clifford Geertz, and "radical empiricism" as stated by the pragmatist philosopher William James.

It is worth noting that during the discussion following the talk, Johanna Drucker proposed to expand this program to other things similar yet distinct from ambiguity: polyvalence, polysemy, multistability, etc.

## 4.13 The Line Graph and the Slave Ship

*Lauren Klein (Emory University – Atlanta, US)*



When we encounter a line graph or a pie chart, we tend to think of the role of visualization – if we think of it at all – as simply revealing the meaning of the data underneath. The reality, however, is that the act of visualizing data generates meaning in and of itself. In this talk, I return to the origins of modern data visualization in order to excavate this meaning. Exploring two examples of early data visualization – the line graphs of British trade data included in William Playfair's Commercial and Political Atlas (1786) and the Diagram of a Slave Ship (1789) created and circulated by a group of British antislavery activists, I connect Enlightenment theories of visual and statistical knowledge to contemporaneous ideas about race and nation. By examining and re-visualizing the data associated with these charts, I will further show how data visualization always carries a set of implicit assumptions – and, at times, explicit arguments – about how knowledge is produced, and who is authorized to produce it. Placing this work in the context of my larger project, Data By Design: An Interactive History of Data Visualization, I will conclude with a consideration of the ethics of visualization in the present. Through a discussion of contemporary examples, I will show how data visualization can bear witness to instances of oppression at the same time that it can – if intentionally designed – hold space for what cannot be conveyed through data alone.

## 4.14 Uncomputational Thinking or What VIS/DH owes to the humanities

*Lamqaddam, Houda (University of Amsterdam, NL)*



In this talk, I discuss the power structures that underlie interdisciplinary work between visualisation and humanities research. Funding cuts in the humanities and the soft power of the computational function to create a power imbalance among collaborators that can lead to "disciplinary capture". I explore three ways that this imbalance affects the research outcomes: through the erasure of humanistic theory, one-sided positioning, and the devaluation of research outcomes. I outline the possible approaches that humanist scan select to engage with visualisation researchers. I end by introducing the concept of "uncomputational thinking" as a method for visualisation research in a humanist context, and point out the importance of slowness and methodological humility in the process.

# 5 Working groups

## 5.1 Complexity

*Derya Akbaba (Linköping University, SE), Alfie Abdul-Rahman (King's College London, GB), Mark-Jan Bludau (FH Potsdam, DE), Michael Correll (Northeastern University – Portland, US), Mennatallah El-Assady (ETH Zürich, CH), Linda Freyberg (DIPF – Berlin, DE), Nicole Hengesbach (University of Warwick – Coventry, GB), Mathieu Jacomy (Aalborg University Copenhagen, DK), Houda Lamqaddam (KU Leuven, BE), Isabel Meirelles (The Ontario College of Art and Design University, CA), Bridget Moynihan (Library and Archives Canada – Ottawa, CA), Fabian Offert (University of California – Santa Barbara, US), Bettina Speckmann (TU Eindhoven, NL), and Florian Windhager (Donau-Universität Krems, AT)*



This breakout group focused on the polysemic concept of *complexity* in visualization. One challenge in visualizing digital humanities data is the complexity of the analyses and the objects of inquiry. For instance, a simple bar chart might be considered sufficient to surface information about a company's sales data over the course of financial quarters (Gelman, 2013), but would not be considered sufficient to capture the important features of a novel corpus or musical scores, or at least not in a way that provides sufficient input for hermeneutical study (Windhager, 2018). There is also a connected tension between the oft-stated goal of visualization to simplify (instantiated in principles like Tufte's maxim to increase the "data-ink ratio" or reduce "chart junk" (Akbaba, 2021) and the hermeneutical impulse to reveal complexity or unpack tacit assumptions. A final related tension was how to manage the complexity "hidden" in superficially simple visualizations (Kostelnick, 2007): for instance, provenance information, methodological choices, and the mathematical sophistication entailed in layout algorithms and dimensionality reduction methods.

The discussions of the group settled on several angles around how complexity can function as a design material. That is, complexity not as an inherently bad or unnecessary component to be reduced as much as possible, but as a material or resource (like color, text, sound, or even machine learning (Dove, 2017; Holmquist, 2017) that can be used to accomplish a number of design goals. Clarifying this concept involved first exploring the different meanings of complexity (Latour, 2008; Latour, 2012; Norman, 2016). For instance, complexity in the size of the data (say, the number of facets or dimensions), complexity of the transformation or representation, visual complexity of the visualization, complexity in the interactive affordances of the visualization, complexity in interpreting the visualization, and the complexity of the communicative goal.

Exploring these different axes of complexity allowed the construction of examples of interesting locations in "complexity space." A scatterplot that directly encodes two quantitative values that are relatively straightforward to describe–say, the number of pages in a novel plotted against the date of publication–is differently complex than a scatterplot of, say, a two dimensional projection of the same novels from a high-dimensional space, even though both result in a very similar, superficially "simple" visualization. The backing data, and what it means to interpret visual features like clusters or outliers, are fundamentally different, even as the objects of inquiry and the visual design are identical.

The conversation then moved to examples of "beneficial" complexity (Hullman, 2011): using complexity to, say, encourage the reader of the visualization to slow down and de-familiarize themselves with the phenomenon of interest (Bradley, 2016), communicate or disclose uncertainty or provenance information, encourage new or serendipitous ways of interacting with the data [McCurdy, 2015; Thudt, 2012) or even to simply communicate that the phenomenon of interest are, in fact, more complex than they might otherwise appear. The group concluded with a (partly provocative) statement: that, as with Tesler's law of UX design (Norman, 2016), there is a "conservation of complexity" in visualization design: a visually simple visualization is likely hiding vast methodological or rhetorical complexity, and a visually complex visualization might allow much more straightforward paths to communication or analysis.

The group explored several potential outputs, settling on an initial multi-faceted study of complexity as design material: collecting interesting examples of artifacts (both visualizations and otherwise) that live in interesting areas of complexity space, definitions and framings of the myriad forms of complexity, and potential functions (both positive uses but also abuses) of complexities in design.

## References


1 Akbaba, D., Wilburn, J., Nance, M.T. and Meyer, M., 2021. *Manifesto for Putting'Chartjunk'in the Trash 2021!* arXiv preprint arXiv:2109.10132.

2 Bradley, A. J., Mehta H., Hancock M. and Collins C. (2016). *Visualization, digital humanities, and the problem of instrumentalism.* Vis4DH 2016.

3 Dove, G., Halskov, K., Forlizzi, J. and Zimmerman, J., 2017. *UX design innovation: Challenges for working with machine learning as a design material.* In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 278-288.

4 Gelman, A., and Unwin A., 2013 *Infovis and statistical graphics: different goals, different looks.* Journal of Computational and Graphical Statistics 22(1), pp. 2-28.

5 Healy, K., 2017. *Fuck nuance.* Sociological Theory 35(2), pp. 118-127.

6 Holmquist, Lars Erik., 2017. *Intelligence on tap: artificial intelligence as a new design material.* Interactions 24(4), pp. 28-33.

7 Hullman, J., Eytan A., and Priti S., 2011. *Benefitting InfoVis with visual difficulties.* IEEE Transactions on Visualization and Computer Graphics 17(12), pp. 2213-2222.

8 Kostelnick, Charles (2007). *The visual rhetoric of data displays: The conundrum of clarity.* IEEE Transactions on Professional Communication 50(4), pp.280-294.

9 Latour, B., 2008. *A cautious Prometheus? A few steps toward a philosophy of design (with special attention to Peter Sloterdijk).* Proceedings of the 2008 annual international conference of the design history society.

10 Latour, Bruno, (2012). *Visualisation and cognition: Drawing things together.* AVANT. Pismo Awangardy Filozoficzno-Naukowej 3, pp. 207-257.

11 McCurdy, N., Lein, J., Coles, K. and Meyer, M., 2015. Poemage: Visualizing the sonic topology of a poem. IEEE transactions on visualization and computer graphics, 22(1), pp.439-448.

**12** Norman, D. A., 2016. *Living with complexity.* MIT press.

**13** Thudt, A., Hinrichs, U., and Carpendale, S., 2012. *The Bohemian Bookshelf: Supporting Serendipitous Book Discoveries through Information Visualization.* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

**14** Windhager, F., Glinka, K., Mayr, E., Schreder, G., Dörk, M. (2018) *Zur Weiterentwicklung des "cognition support": Sammlungsvisualisierungen als Austragungsort kritisch-kulturwissenschaftlicher Forschung.* Digital Humanities im Deutschsprachigen Raum (DHd).


## 5.2 Encodings

*Richard Brath (Uncharted Software – Toronto, CA), Alfie Abdul-Rahman (King's College London, GB), Charles Berret (Linköping University, SE), Mark-Jan Bludau (FH Potsdam, DE), Johanna Drucker (University of California at Los Angeles, US), Aida Horaniet Ibanez (University of Luxembourg, LU), Johannes Liem (Donau-Universität Krems, AT), Fabian Offert (University of California – Santa Barbara, US), and Christophe Schuwey (Université de Bretagne Sud, FR)*



Many predefined data types exist (time, hierarchy, connections, geography, ordered, etc.), each with inherent visual representations. Each of these data characterizations and implicit encodings limit interpretation. Where is materiality, experience, missing data, contemporary context, and so on?

If current visualization techniques are understood to be "statistically-based visual data analysis", then there are many gaps to humanities needs, which include:

- no given data schema
- completely flexible encoding environment
- support for multiple simultaneous hypotheses
- abilities to fragment and recombine data
- easy access to sources and context
- high multi-dimensionality but not dimensionality reduction
- no singular narrative: many, but also narratives of mis- and dis-information
- uncertainty is not simplistic: potentially thousands of aspects of uncertainty

Furthermore, the objectives of the current standard approach to data visualization include fast and unambiguous decoding of representations. However, slow reading is an explicit objective for the humanities. Many encodings dissuaded or underexplored should be re-examined: 3D, imagery, glyphs, etc... 3D visualization, for example, requires a point of view, there is no singular god-like overview with 3D. Humanities needs a wider, flexible graphical vocabulary, and per project, or per analysis. Humanities visualization needs to embrace the complexity of slow reading (Brath, 2023).

The act of creating and generating a visualization should be an interpretative process: sketching supports this well, whereas data visualization tools require training which is too hard. The current approach to data visualization pre-supposes data: the broader situational context includes phenomena > corpora > data > sketches > encodings > uncertainty > visualization, which can be cycled through in any order.

One effective direction of future research is a combination of hand-drawn sketching of visualizations with simple interactions to converse with data and iterate: for example, I can connect my data to the sketch, or use the sketch to search and extract data, or use difficulties within this conversation to prompt new avenues of exploration and interpretation (possibly using complementary AI).

The discussion has shown a lot of avenues for theoretical and practical explorations that could culminate in a potential anthology that includes provocations and speculative case studies around the issues around encodings and visualization in the humanities.

### References

1 Brath, R., 2023. *Visualizing humanities data... slowly.* https://richardbrath.wordpress.com/2023/10/13/visualizing-humanities-data-slowly/

## 5.3 Vis in the Humanities & Encodings (output session)

*Richard Brath (Uncharted Software – Toronto, CA), Johanna Drucker (University of California at Los Angeles, US), Johannes Liem (Donau-Universität Krems, AT), Christophe Schuwey (Université de Bretagne Sud, FR), and Joris van Zundert (Huygens Institute – Amsterdam, NL)*



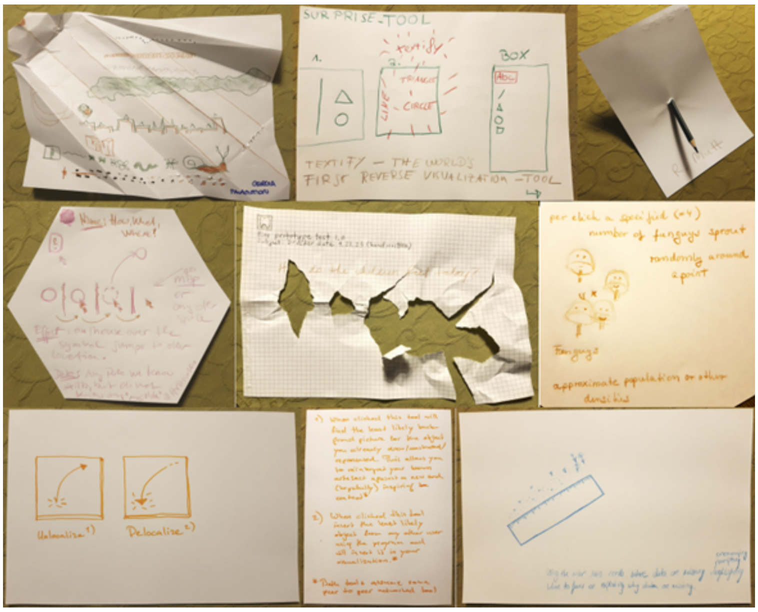
The discussion group *Vis in the Humanities* & *Encodings'* addressed the question of the need for novel visualization techniques in the humanities. Humanistic interpretive activity does not necessarily fit well with traditional computational quantitative visualization nor a linear method sequencing hypothesis, data, development, and evaluation. Furthermore, current visualization does not work well with heterogeneous data nor polyvalent interpretation.

At the highest level, group members felt that the issue could be best addressed through a prompt for speculative case studies, and reference to case studies related to humanities visualization but not perceived as such. A few examples include:

- Temporality: e.g. the decay of Damien Hurst's shark
- Perception of data: What lens or perspective is used? For example, currency has no intrinsic value outside the conditions of which value is assessed.
- Instantiation transforms interpretation: e.g. Shakespeare's Sonnet 18
- Illusion of stability: works are continuous re-interpreted, e.g. fan fiction

To test the provocations, we directly engaged 30 workshop participants with the following prompt: *"Remember when paint programs had cool brushes that create wacky effects? Sketch a tool that has unexpected behaviors. Make a visual effect. Then imagine what kind of data it could represent."* After 20 minutes of drawing 32 responses were assembled (see below).

Individuals selected a drawing at random, presented it to the group, with a brief community interpretation. Some representative examples are shown below. Provocation outputs generated novel visualization transformations, such as reversing visual encodings into text, automated contextualization, exposure of data gaps with biological generative data fills, interpretation resistance encoding and so on. There are a number of publications feasible for the outputs, as well as a reusable method.




## References


1 Hurst, D., 1991. The Physical Impossibility of Death in the Mind of Someone Living https://en.wikipedia.org/wiki/The_Physical_Impossibility_of_Death_in_the_Mind_of_Someone_Living

## 5.4 Historical examples – of VIS4DH

*Richard Brath (Uncharted Software – Toronto, CA), Derya Akbaba (Linköping University, SE), Jeremy Douglass (University of California – Santa Barbara, US), Johanna Drucker (University of California at Los Angeles, US), Mennatallah El-Assady (ETH Zürich, CH), and Florian Windhager (Donau-Universität Krems, AT)*



The key outcome from this working group is expected to be an open collection of visualizations sourced from the humanities and from the history of using graphical representations of data in this field. This is required because a rather small set of historical examples currently perpetuates that "visualization must look like these". Related challenges include

- The collection of related meta-data, forsuch as source, provenance, why it is a visualization, contents description, medium, date, authorship, copyright permission, interpretation, references (e.g. grant proposals), and related files (e.g. dataset, URL, etc).
- Infrastructure to collect and organize the visualizations.
- Some form of definition of "visualization" and "history".
- The need to integrate variety of existing sources of historic data visualizations (although not necessarily humanities-centric), including for example the Data Visualization Society's slack channel (vizsociety.slack.com) and contributions tagged #topic-historical-viz.

## 5.5 What is / How are data?

*Richard Brath (Uncharted Software – Toronto, CA), Johanna Drucker (University of California at Los Angeles, US), Yanni Loukissas (Georgia Institute of Technology – Atlanta, US), Isabel Meirelles (The Ontario College of Art and Design University, CA), and Fabian Offert (University of California – Santa Barbara, US)*

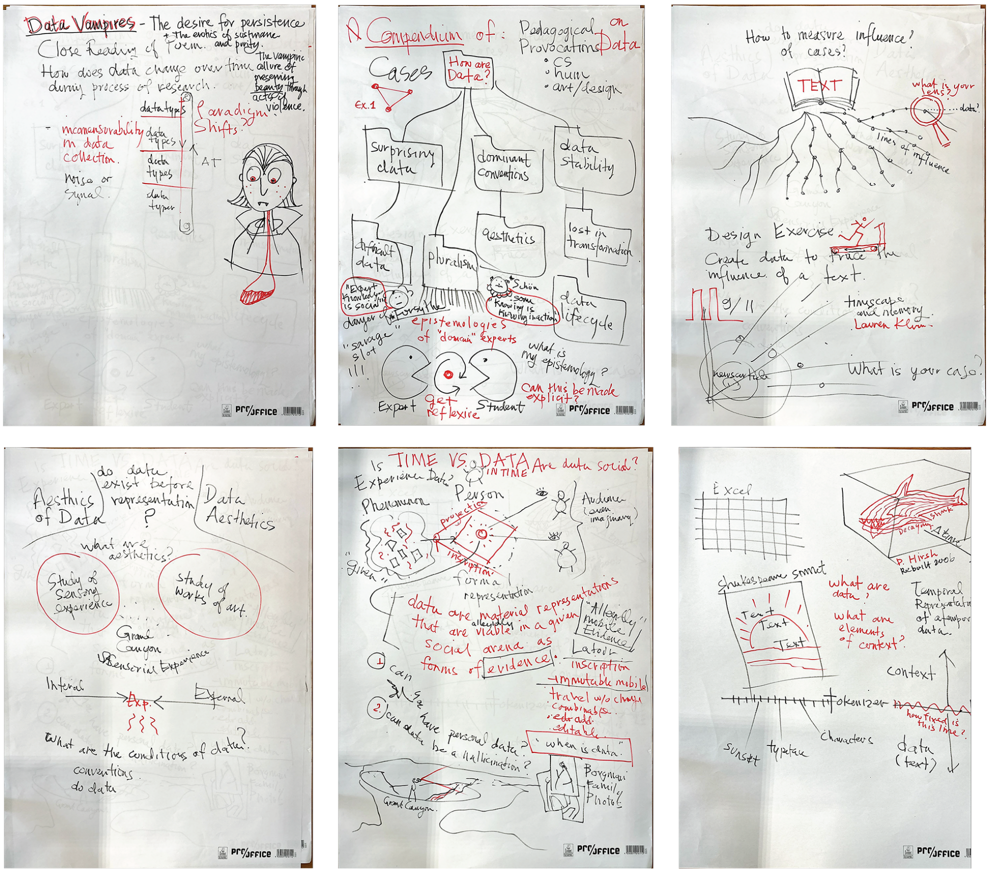


*What is data?* Arguably, for VIS4DH we need more pluralistic approaches with tolerance for different epistemologies. Positivist work needs to acknowledge its framework – as should all other approaches towards "drawing things together" (Latour, 1999; 2011).[2] This could be an essential contribution which humanities can offer to the field of visualization. From this working group's point of view, it is a matter of methodological rigor to talk about *what you mean by data* – and a range of closely related quetions – when we build our work on this concept.

- How to make explicit what we mean by data? (Expansive, inclusive, but explicit)
- What is data? (E.g., a mix of affective information)
- Where does data stop and context start?

[2] Latour's account of sampling soil through overlay to abstract mathematical representation. Gaps between can't be bridged, they are always jumped. Same thing happens on the mathematical side – throw away something. Collect the data outliers? The stack? What is thrown away at each level of the stack? Scientific images allow you to reason in ways that are different from text. Bucket of contingencies. Decisions about removing/simplifying are meant to assist cognitive processing.

- How to bring in the context? (Discrete numbers vs. what context)
- How is data?
- What happens before the spread sheet?
- What is the relation between unstructured data and structured data?
- Unstructured vs. unstructurable? (Offenhuber, 2019).
- What data does not lend itself to visualization – or can all data be visualized?
- What are the instruments used to collect the data? (Data lifecycles are worth capturing)




Further discussions of this working group revolved around the relationship between *data and aesthetics* as motivated by the questionable criticism of data being "aestheticized" as a threat to objectivity and purity of data. Aesthetic judgments and categories in the world include formal characteristics and ways of looking at and talking about them (e.g. "cute") that register consumer categories. Criticism of this sort appears not as valuable as it claims, synchronously over-producing and under-producing quality criteria. However, what is the relationship between data and "cute", "gimmicky", "zany", etc.? Could we think about data as an aesthetic category, which is necessarily culturally situated, embedded, then expressed and instantiated? In order to be recognizable, data has to be a rhetorical category and should count as evidence for some claim you are making. *Data is what counts* – not in the quantitative sense but in the relevance – so lay them out on a page and then define systems of measure. Be explicit about what "counts". Data as an aspect of what you want to know, desire?

During their discussions, the group deliberately aimed for documenting their discourse in a multimodal fashion, as illustrated by Yanni Loukissas' graphic recordings of related conceptual and practical challenges (figure below).

**Working group's action points.**

- A book of provocations.
- Example of things that are data in a surprising way.
- What counts as data? A typology of dominant cultural forms across domains.
- Difficult data – a book of this.

**References**


1 Galloway, A., 2011. Are some things unrepresentable? Theory, Culture & Society, 28(7-8), pp.85-102.
2 Latour, B., 2011. Drawing things together. The map reader: Theories of mapping practice and cartographic representation, pp.65-72
3 Latour, B., 1999. Pandora's hope: Essays on the reality of science studies. Harvard university press.
4 Offenhuber, D., 2019. Data by proxy – material traces as autographic visualizations. IEEE transactions on visualization and computer graphics, 26(1), pp.98-108.
5 Risam, R., 2019. Beyond the migrant "problem": Visualizing global migration. Television & New Media, 20(6), pp.566-580.


## 5.6 Visualization Literacy


*Aida Horaniet Ibanez (University of Luxembourg, LU), Derya Akbaba (Linköping University, SE), Alejandro Benito-Santos (University of Salamanca, ES), Jeremy Douglass (University of California – Santa Barbara, US), Jean-Daniel Fekete (INRIA Saclay – Orsay, FR), Jan Horstmann (Universität Münster, DE), Mathieu Jacomy (Aalborg University Copenhagen, DK), Steffen Koch (Universität Stuttgart, DE), Yanni Loukissas (Georgia Institute of Technology – Atlanta, US), Isabel Meirelles (The Ontario College of Art and Design University, CA), David Pao (Royal College of Art – London, GB), and Florian Windhager (Donau-Universität Krems, AT)*




During the two breakout group sessions on literacy, discussions revolved around the central term (and consequently the scope), i.e., whether it was "data literacy", "data visualization literacy" or "visualization literacy". It was agreed that the most appropriate (and inclusive) term was "visualization literacy". Major concerns about the topic included the limitation of the current definition to the understanding of statistical graphs (e.g., bar charts, scatter plots) and networks, the lack of agreement on what other methods and techniques should be considered under the concept of "visualization literacy", the lack of teaching materials including feedback from instructors, and the difficulty in addressing students with different backgrounds.

Understanding "literacy" exclusively as "statistical charts literacy" leads to teaching a limited number of charts and tools, that promote summarization and reduction for the rapid consumption of information, a necessary aspect in some analyses, but limiting in others, especially in the humanities. At the same time, a standardized visual vocabulary (e.g., bars, lines, grids) does not allow other encoding options required for rhetorical expression and interpretation. This led to the question: what else should be included in the concept of "visualization literacy"? Among others, we discussed: the use of free encodings adapted to each visualization as proposed in data humanism (and consequently slow reading), the use of visual vocabularies for interpretation (e.g., repulsion, impact, fold), the study of non-representational approaches (i.e., starting with the visualization), the visualization of different temporalities, and the analysis of comics using visualization.

The teaching materials available are limited in content, as described above, and are not always adapted to the different contexts. This creates a barrier for many and makes it especially difficult to teach groups with students with different prior knowledge (e.g., computer science vs. humanities training), a common scenario in the digital humanities.

As future steps, it was decided to create a more comprehensive repository of teaching materials, which will also collect detailed feedback from users. We also set out to write a manifesto on the need to expand the concept of "visualization literacy", and therefore its teaching, evaluation, and impact on the development and use of tools.

## 5.7 Theories & Methods

*Rabea Kleymann (TU Chemnitz, DE), Alejandro Benito-Santos (University of Salamanca, ES), Stefania Forlini (University of Calgary, CA), Linda Freyberg (DIPF – Berlin, DE), Uta Hinrichs (University of Edinburgh, GB), Lauren Klein (Emory University – Atlanta, US), Yanni Loukissas (Georgia Institute of Technology – Atlanta, US), Bridget Moynihan (Library and Archives Canada – Ottawa, CA), Joris van Zundert (Huygens Institute – Amsterdam, NL), and Florian Windhager (Donau-Universität Krems, AT)*



The discussion group on *Theories* delved into the question and role of theories in describing the relationship between the two research communities of visualization (VIS) and the (digital) humanities. The concept of theory is fundamental to research. Theories in any form reflect the conditions for the possibility of building knowledge and are integral to research settings. Within the discussion, two central approaches were delineated. There was an in-depth exploration of the understanding of the term "theory."

- What is understood by theory?
- What role do theoretical considerations play in everyday research and collaborations?
- How do theories differ from methods?

Against this backdrop, we engaged deeply in a survey on relevant research methods, theories, and approaches in the field of VIS+DH which has been conducted prior to the seminar, and with which participants documented their guiding concepts and procedures with short fact sheets (see Figure).

Building on the repository of theoretical approaches gathered, we then turned our attention to theory-specific implications, promises and pitfalls for interpreting visualizations, designing visualizations, and research practices in both communities. In this context, we worked through three exemplary theoretical schools – Hermeneutics, Semiotics, and Critical Theory – to explore what it might mean to (re)view visualization as an outcome and process through


these theoretical lenses. What becomes visible and negotiable when we create and interpret visualizations hermeneutically, semiotically, or critically? How do theoretical considerations manifest in visualizations?

One major outcome of the discussion was the observation that the theoretical pluralism in the humanities provides an opportunity to negotiate (data) visualizations in their contingency and situatedness. As an initial output, we aim for an article that focuses on the multiperspective potential of theories and visualizations. Inspired by Steven Wallace's poem "Thirteen Ways of Looking at a Blackbird" (1954), the idea is that theories allow us to illuminate different aspects of a data set and/or data visualizations. Another associated output is a visual glossary of theory for/in visualization, representing a kind of "theory browser".

## References

1 Wallace, S., 1954, *Thirteen Ways of Looking at a Blackbird*. https://www.poetryfoundation.org/poems/45236/thirteen-ways-of-looking-at-a-blackbird

## 5.8 Ethics

*Georgia Panagiotidou (King's College London, GB), Alfie Abdul-Rahman (King's College London, GB), Michael Correll (Northeastern University – Portland, US), Leonardo Impett (University of Cambridge, GB), Lauren Klein (Emory University – Atlanta, US), and Geoffrey Rockwell (University of Alberta – Edmonton, CA)*



As information visualizations are increasingly used to engage citizens on social and political issues, this interdisciplinary group came together to discuss the *ethics of visualization*. The group started broadly, by discussing what is ethical, and quickly came to a discussion of case studies, each from a different discipline and/or perspective. We talked about the neutrality of data, appropriation of voice, neutrality of the visualization in itself as a perceptual, cultural and historical object as well as an interface with issues such as accessibility and interactivity.

The group decided to take action on these discussions by setting up a series of case studies of varied nature that illustrate the ethical entanglements of data and their visualization. As a preliminary step towards this case study synthesis, the group has been holding monthly meetings since the original Dagstuhl Seminar and submitted a short paper to the ADHO conference exposing their approach and initial sketches for these case studies. The paper initiates a discussion on the ethics of visualization through the development of six case studies ranging from historical examples of slave trade and anatomy to current day pandemic, climate, and algorithmically designed visualizations. These case studies are meant to form an initial part of a reflective educational activity for visualization students, designers, and practitioners.

### References

1. Collaborative Indigenous Research (n.d.) Collaborative Indigenous Research. Accessed 11 December 2023. https://www.collaborativeindigenousresearch.com.
2. Data Journalism Handbook 2 (n.d.) DataJournalism.com https://datajournalism.com/readhandbook/two
3. The First Nations Principles of OCAP (n.d.) The First Nations Information Governance Centre. Accessed 11 December 2023.https://fnigc.ca/ocap-training/.
4. Nations to Nations: Indigenous Voices at Library and Archives Canada. (n.d.) Accessed 11 December 2023. https://library-archives.canada.ca/eng/collection/engage-learn/publications/ebooks/Pages/nations-to-nations-indigenous-voices-lac.aspx.
5. Cairo, A., 2019. How charts lie: Getting smarter about visual information. WW Norton & Company.
6. D'Ignazio, C. and Klein, L.F., 2023. Data feminism. MIT press.
7. Estill, L., 2019. "Digital Humanities' Shakespeare Problem" Humanities 8, no. 1: 45. https://doi.org/10.3390/h8010045
8. Tuck, E., and Yang, K. W., 2014. Unbecoming claims: Pedagogies of refusal in qualitative research. Qualitative Inquiry, 20(6), 811-818.
9. Tuck, E. & Yang, K.W., 2021. Decolonization is not a metaphor. Tabula Rasa, (38), pp.61-111.

## 5.9 Surveying the State-of-the-Art

*Alejandro Benito-Santos (University of Salamanca, ES), Alfie Abdul-Rahman (King's College London, GB), Jean-Daniel Fekete (INRIA Saclay – Orsay, FR), Aida Horaniet Ibanez (University of Luxembourg, LU), Lauren Klein (Emory University – Atlanta, US), Rabea Kleymann (TU Chemnitz, DE), and Florian Windhager (Donau-Universität Krems, AT)*



This breakout group deliberated on research directions for a survey titled *Visualization for the Digital Humanities.* The initial discussion revolved around defining the survey's scope and boundaries (McNabb, 2019), pinpointing pertinent research questions, and enhancing these through insights from recent, related surveys (Benito-Santos, 2020a; Benito-Santos, 2020b; Jänicke, 2015; Jänicke, 2017; Windhager, 2018). A pivotal part of the dialogue focused on the prevailing types of visualizations in digital humanities, their impact, and the chronological development of visualization tools. A critical objective of the survey is to pinpoint significant "inflection points" – such as influential papers, books, libraries, and tools – that have markedly shaped the field. The group highlighted two notable gaps in existing literature: The need for an in-depth understanding of how current visualization systems foster diverse interpretations and perspectives on data, and a lack of insight into the evaluation techniques (Isenberg, 2013) employed in VIS4DH practice. Addressing these gaps, the survey aims to encapsulate current trends, identify potential deficiencies, and spotlight underutilized techniques, thereby providing comprehensive guidance for future research in this area.

The consensus was that the survey should encompass a wide range, covering research articles in visualization and humanities across various data types, including texts and images. It should emphasize recent developments, best practices, and seminal publications. The intended audience spans both humanities and visualization practitioners, suggesting a potential dual publication approach. The data collection methodology will integrate quantitative scraping with qualitative analysis, underscoring the significance of datasets, theories, methods, and evaluation techniques, thus ensuring a well-rounded and impactful survey.

### References

1 McNabb, L., and Laramee, R.S., 2019. *How to Write a Visualization Survey Paper: A Starting Point*, in: Tarini, M., Galin, E. (Eds.), Eurographics 2019 – Education Papers. The Eurographics Association. https://doi.org/10.2312/eged.20191026

2 Benito-Santos, A., and Therón Sánchez, R., 2020a. *A Data-Driven Introduction to Authors, Readings, and Techniques in Visualization for the Digital Humanities.* IEEE Computer Graphics and Applications 40, 45–57. https://doi.org/10.1109/MCG.2020.2973945

3 Benito-Santos, A., and Therón, R., 2020b. *Pilaster: A Collection of Citation Metadata Extracted From Publications on Visualization for the Digital Humanities*, in: Proceedings of the 2020 IEEE 5th Workshop on Visualization for the Digital Humanities (VIS4DH) , pp. 24–29. https://doi.org/10.1109/VIS4DH51463.2020.00009

4 Jänicke, S., Franzini, G., Cheema, M.F., and Scheuermann, G., 2017. *Visual Text Analysis in Digital Humanities.* Computer Graphics Forum 36, 226–250. https://doi.org/10.1111/cgf.12873

5 Jänicke, S., Franzini, G., Cheema, M.F., and Scheuermann, G., 2015. *On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges*, in: Borgo, R., Ganovelli,

F., Viola, I. (Eds.), Eurographics Conference on Visualization (EuroVis) – STARs. The Eurographics Association. https://doi.org/10.2312/eurovisstar.20151113

6 Windhager, F., Federico, P., Schreder, G., Glinka, K., Dörk, M., Miksch, S., and Mayr, E., 2018. *Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges.* IEEE Transactions on Visualization and Computer Graphics 1–1. https://doi.org/10.1109/TVCG.2018.2830759

7 Isenberg, T., Isenberg, P., Chen, J., Sedlmair, M., and Möller, T., 2013. *A Systematic Review on the Practice of Evaluating Visualization.* IEEE Transactions on Visualization and Computer Graphics 19, 2818–2827. https://doi.org/10.1109/TVCG.2013.126

## Participants

- Alfie Abdul-Rahman
King's College London, GB
- Derya Akbaba
Linköping University, SE
- Alejandro Benito-Santos
University of Salamanca, ES
- Charles Berret
Linköping University, SE
- Mark-Jan Bludau
FH Potsdam, DE
- Richard Brath
Uncharted Software – Toronto, CA
- Michael Correll
Northeastern University – Portland, US
- Jeremy Douglass
University of California – Santa Barbara, US
- Johanna Drucker
University of California at Los Angeles, US
- Mennatallah El-Assady
ETH Zürich, CH
- Jean-Daniel Fekete
INRIA Saclay – Orsay, FR
- Stefania Forlini
University of Calgary, CA
- Linda Freyberg
DIPF – Berlin, DE
- Katrin Glinka
Freie Universität Berlin, DE
- Nicole Hengesbach
University of Warwick – Coventry, GB
- Uta Hinrichs
University of Edinburgh, GB
- Aida Horaniet Ibanez
University of Luxembourg, LU
- Jan Horstmann
Universität Münster, DE
- Leonardo Impett
University of Cambridge, GB
- Mathieu Jacomy
Aalborg University Copenhagen, DK
- Lauren Klein
Emory University – Atlanta, US
- Rabea Kleymann
TU Chemnitz, DE
- Steffen Koch
Universität Stuttgart, DE
- Houda Lamqaddam
KU Leuven, BE
- Johannes Liem
Donau-Universität Krems, AT
- Yanni Loukissas
Georgia Institute of Technology – Atlanta, US
- Isabel Meirelles
The Ontario College of Art and Design University, CA
- Bridget Moynihan
Library and Archives Canada – Ottawa, CA
- Fabian Offert
University of California – Santa Barbara, US
- Georgia Panagiotidou
King's College London, GB
- David Pao
Royal College of Art – London, GB
- Geoffrey Rockwell
University of Alberta – Edmonton, CA
- Christophe Schuwey
Université de Bretagne Sud, FR
- Bettina Speckmann
TU Eindhoven, NL
- Joris van Zundert
Huygens Institute – Amsterdam, NL
- Florian Windhager
Donau-Universität Krems, AT

Report from Dagstuhl Seminar 23391

# The Futures of Reactive Synthesis

**Nathanaël Fijalkow*[1], Bernd Finkbeiner*[2], Guillermo A. Pérez*[3], Elizabeth Polgreen*[4], and Rémi Morvan†[5]**

1 **CNRS – Talence, FR.** nathanael.fijalkow@gmail.com
2 **CISPA – Saarbrücken, DE.** finkbeiner@react.uni-saarland.de
3 **University of Antwerp, BE.** guillermoalberto.perez@uantwerpen.be
4 **University of Edinburgh, GB.** elizabeth.polgreen@ed.ac.uk
5 **University of Bordeaux, FR.** remi.morvan@u-bordeaux.fr

**Abstract**

The Dagstuhl Seminar 23391 "The Futures of Reactive Synthesis" held in September 2023 was meant to gather neighbouring communities on a joint goal: Reactive Synthesis. We identified five trends: neural-symbolic computation, template-based solving for constraint programming, symbolic algorithms, syntax-guided synthesis, and model learning; and the objective was to discuss the potential futures of the field.

**Seminar** September 24–29, 2023 – https://www.dagstuhl.de/23391

**2012 ACM Subject Classification** Computing methodologies → Artificial intelligence; Theory of computation → Formal languages and automata theory; Computing methodologies → Parallel programming languages

**Keywords and phrases** program synthesis, program verification, reactive synthesis, temporal synthesis

**Digital Object Identifier** 10.4230/DagRep.13.9.166

## 1 Executive Summary

*Nathanaël Fijalkow*
*Bernd Finkbeiner*
*Guillermo A. Pérez*
*Elizabeth Polgreen*



This report documents the program and the outcomes of Dagstuhl Seminar 23391 "The Futures of Reactive Synthesis".

The seminar was meant to gather neighbouring communities on a joint goal: Reactive Synthesis. We identified five trends: neural-symbolic computation, template-based solving for constraint programming, symbolic algorithms, syntax-guided synthesis, and model learning. They were represented by different participants, and in particular by four invited speakers. We had three female invited speakers and one male invited speaker; all delivered very insightful and forward-thinking talks:

- Anne-Kathrin Schmuck
- Armando Solar-Lezama
- Ruzica Piskac
- Dana Fisman

* Editor / Organizer
† Editorial Assistant / Collector

We introduced a number of mechanisms to encourage discussions and the exchange of ideas: an open problem session, long Q&A sessions after each invited talk, and most importantly working sessions. The working sessions were proposed by participants (who volunteered in advance, after a call by email to all participants). The proposer would have a few minutes to introduce the topic they would like to discuss. Each session included 3 or 4 different topics, discussed in parallel in smaller groups. In each case, we had (by some miracle!) a fair division of all participants into the 3 or 4 topics, and we had very good feedback that many working sessions resulted in very fruitful and insightful discussions. We had "progress report sessions" where the leaders of the working sessions gave a 5 or 10-min summary of the discussions.

We also had 9 contributed talks from participants, responding to an open call. They were 20 minutes each, and greatly contributed to getting all participants involved and for representing all trends and recent advances in the field.

We as organizers had very good feedback about the organization of the week: the rather light schedule gave enough time for people to discuss, and the different talks and organized sessions gave enough ways to get to know new people and topics. The seminar included a number of junior participants, who got to meet experts in the field. The mix of tools and theory topics covered during the seminar gives us hope that it will yield results both in the short and long term.

## 2 Table of Contents

# 3 Overview of Talks

## 3.1 Solving Infinite-State Games via Acceleration

*Rayna Dimitrova (CISPA – Saarbrücken, DE)*


**Joint work of** Philippe Heim, Rayna Dimitrova
**Main reference** Philippe Heim, Rayna Dimitrova: "Solving Infinite-State Games via Acceleration", CoRR, Vol. abs/2305.16118, 2023.
**URL** https://doi.org//10.48550/ARXIV.2305.16118

Two-player graph games have found numerous applications, most notably in the synthesis of reactive systems from temporal specifications, but also in verification. The relevance of infinite-state systems in these areas has lead to significant attention towards developing techniques for solving infinite-state games. In this talk I will present novel symbolic semi-algorithms for solving infinite-state games with omega-regular winning conditions. The novelty of our approach lies in the introduction of an acceleration technique that enhances fixpoint-based game-solving methods and helps to avoid divergence. Classical fixpoint-based algorithms, when applied to infinite-state games, are bound to diverge in many cases, since they iteratively compute the set of states from which one player has a winning strategy. Our proposed approach can lead to convergence in cases where existing algorithms require an infinite number of iterations. This is achieved by acceleration: computing an infinite set of states from which a simpler sub-strategy can be iterated an unbounded number of times in order to win the game. Ours is the first method for solving infinite-state games to employ acceleration. Thanks to this, it is able to outperform state-of-the-art techniques on a range of benchmarks, as evidenced by our evaluation of a prototype implementation.

## 3.2 Fixpoint Equations for Synthesis – Towards a Renewed Interest

*Rüdiger Ehlers (TU Clausthal, DE)*


**Joint work of** Ayrat Khalimov, Rüdiger Ehlers

Reactive synthesis is traditionally reduced to solving a game between two players, where the game graph and the winning condition for one of the players in the game encodes the specification and the known information about the environment of the system to be synthesized. In this context, it is customary to encode all the available information into the game graph itself, so that a simple parity, Rabin, or Streett winning condition (among others) remains to be applied to the game graph. This allows to use game solving algorithms optimized for the respective winning condition off-the-shelf to solve the synthesis games.

Combining complex game graphs and relatively simple winning conditions is however not the only way to approach game-based reactive synthesis. We can alternatively distil only a part of the available information (such as the known information about the environment of the system to be synthesized and some simple specification parts) into the game graph, and encode the more complicated specification parts into the winning condition. In practice, this means computing a fixpoint equation that is evaluated over the game graph, where the result of evaluating the equation is the set of game positions from which the specification is realizable. This approach is followed in the Generalized Reactivity(1) Synthesis algorithm by

Piterman, Pnueli, and Sa'ar, which exploits the fact that for the specifications supported for it, there is a simple fixpoint formula template that can be instantiated for any specification of the supported specification class. In this way, the fixpoint equation can be evaluated symbolically if the game graph is easy to encode symbolically, which helped with scaling Generalized Reactivity(1) synthesis to a good number of applications in robotics and control.

In this talk, we discuss one commonly known, one recent, and one new result on computing fixpoint equations encoding complex specifications that go beyond Generalized Reactivity(1) synthesis. All the discussed results are applicable to symbolically represented game graphs. Apart from reviewing how to build such fixpoint formulas from deterministic parity automata for a given specification to be enforced in a game graph, we discuss a recent result by Hausmann, Lehaut, and Piterman and give a summary of our own results on translating a polynomial-time minimizable chain-of-co-Büchi-automata representation of a given omega-regular specification to a fixpoint equation. We provide some experimental results and employ them to argue for establishing a branch of reactive synthesis research that aims at computing efficient to evaluate fixpoint equations over symbolic game graphs. Focusing on such fixpoint equations has three advantages: Firstly, even at the current early state of research, the first approaches are already faster than previous full-LTL synthesis tools on specifications that decompose quite naturally into a game graph and a complex specification. Then, a compilation of a reactive synthesis problem to a game graph plus a fixpoint formula is a concise starting point for performing symbolic reasoning beyond the use of BDDs. Finally, fixpoint equations encoding complex specifications for synthesis would be useful for tackling the synthesis problem in implicitly represented infinite state spaces of games, which may be interesting for control and robotics applications.

This talk led to a working session.

## 3.3 Synthesis from LTL specifications and examples

*Emmanuel Filiot (UL – Brussels, BE)*


**Joint work of** Emmanuel Filiot, Mrudula Balachander, Jean-François Raskin

We study a variant of the problem of synthesizing Mealy machines that enforce LTL specifications against all possible behaviours of the environment including hostile ones. In the variant studied here, the user provides the high level LTL specification $S$ of the system to design, and a set $E$ of examples of executions that the solution must produce. The examples are used to guide the synthesis procedure, and are generalized as much as possible, while preserving realizability of the specification. This talk presents some approach to this problem based on a combination of RPNI automata learning and antichain-based LTL synthesis methods.

### References

1 Mrudula Balachander, Emmanuel Filiot and Jean-François Raskin. *LTL Reactive Synthesis with a Few Hints*. Tools and Algorithms for the Construction and Analysis of Systems (TACAS), 2023.

## 3.4 A primer on reactive synthesis

*Bernd Finkbeiner (CISPA – Saarbrücken, DE)*



The synthesis of reactive systems has been actively investigated since the inception of the problem by Alonzo Church more than sixty years ago. This talk gives an overview of the main results of the area and an outlook on potential future directions to be discussed in the seminar, including neural-symbolic computation and, more generally, machine learning techniques, template-based solving in the context of constraint programming, active learning algorithms, and connections to program synthesis and in particular Syntax Guided Synthesis.

## 3.5 $\omega$-Automata Learning

*Dana Fisman (Ben Gurion University – Beer Sheva, IL)*


Joint work of Dana Angluin, Timos Antonopoulos, Udi Boker, Dana Fisman, Nevin George, Yaara Shoval

This talk surveys the results on learning automata models for regular languages of infinite words. It discusses several positive and negative results across different learning paradigms. The positive results are mostly for automata models that are less common, in particular families of DFAs (FDFAs), strongly unambiguous Büchi automata (SUBAs) and mod-2 multiplicity automata (M2MA). These models have other good qualities, in particular the complexity of the boolean operations (intersection, union, complementation) and decision problems (emptiness, inclusion, equivalence) are good compared to the common omega-automata types. It is thus worth exploring whether they can also be usable for model checking and synthesis of reactive systems.

**References**

1. Dana Angluin, Dana Fisman *Learning regular omega languages*. Theor. Comput. Sci. 650: 57-72 (2016)
2. Dana Angluin, Udi Boker, Dana Fisman *Families of DFAs as Acceptors of $\omega$-Regular Languages*. Log. Methods Comput. Sci. 14(1) (2018)
3. Dana Angluin, Timos Antonopoulos, Dana Fisman *Strongly Unambiguous Büchi Automata Are Polynomially Predictable With Membership Queries*. CSL 2020: 8:1-8:17
4. Dana Angluin, Timos Antonopoulos, Dana Fisman, Nevin George *Representing Regular Languages of Infinite Words Using Mod 2 Multiplicity Automata*. FoSSaCS 2022: 1-20
5. *Constructing Concise Characteristic Samples for Acceptors of Omega Regular Languages*.

## 3.6 Compositional Synthesis with Hyperproperties

*Niklas Metzger (CISPA – Saarbrücken, DE)*



The distributed synthesis problem is to translate a logical specification of a distributed system into an implementation that is guaranteed to satisfy the specification. What makes the synthesis of distributed systems far more challenging than standard reactive synthesis is that each component only has partial knowledge of the global system state. Currently, there are no scalable algorithms for distributed synthesis. The challenge is to devise a compositional synthesis method, i.e., a method that constructs one component at a time. The fundamental difficulty is that the components often need to act upon information that is available only in another component. However, we do not know how that component encodes the information before we know its implementation; seemingly, it is impossible to build one component without knowing the implementation of the other. In this talk, I will present a compositional synthesis method based on the key idea of characterizing the necessary flow of information between the components as a hyperproperty. We introduce information flow assumptions, which are requirements that are necessary in order to realize a particular component. By formulating these assumptions as hyperproperties, we avoid referring to any particular encoding of the information. We develop methods that automatically derive information flow assumptions from the specification and a technique for the automatic synthesis of component implementations based on information flow assumptions. Together, these methods provide a compositional approach to the synthesis of distributed systems.

## 3.7 Ups and downs of distributed synthesis

*Anca Muscholl (University of Bordeaux, FR)*


**Joint work of** Hugo Gimbert, Corto Mascle, Anca Muscholl, Igor Walukiewicz
**Main reference** Hugo Gimbert, Corto Mascle, Anca Muscholl, Igor Walukiewicz: "Distributed controller synthesis for deadlock avoidance", CoRR, Vol. abs/2204.12409, 2022.
**URL** https://doi.org//10.48550/ARXIV.2204.12409

The talk gave an overview of several approaches to distributed reactive synthesis: Pnueli & Rosner model, controller synthesis for Zielonka automata and controller synthesis for lock-sharing systems. While partial information is a direct source of undecidability in the Pnueli & Rosner model, full causal information does not guarantee decidability either (cf. Gimbert 2022). Loose synchronization as in lock-sharing systems allows to recover decidability of controller synthesis at reasonable cost.

## 3.8 Making New Friends in Software Synthesis

*Ruzica Piskac (Yale University – New Haven, US)*


**Joint work of** Wonhyuk Choi, Bernd Finkbeiner, Ruzica Piskac, Mark Santolucito, Felix Klein
**Main reference** Wonhyuk Choi, Bernd Finkbeiner, Ruzica Piskac, Mark Santolucito: "Can reactive synthesis and syntax-guided synthesis be friends?", in Proc. of the PLDI '22: 43rd ACM SIGPLAN International Conference on Programming Language Design and Implementation, San Diego, CA, USA, June 13 – 17, 2022, pp. 229–243, ACM, 2022.
**URL** https://doi.org//10.1145/3519939.3523429

While reactive synthesis and syntax-guided synthesis (SyGuS) have seen enormous progress in recent years, combining the two approaches has remained a challenge. To overcome this obstacle, we introduced Temporal Stream Logic (TSL) [1], a new temporal logic that separates control and data. We developed a CEGAR-like synthesis approach for the construction of implementations that are guaranteed to satisfy a TSL specification for all possible instantiations of the data processing functions. However, specifications often involve interpreted functions: for example, arithmetic functions or string manipulations. We extended TSL to Temporal Stream Logic modulo theories (TSL-MT) [2], a framework that unites the two approaches to synthesize a single program. In our approach, reactive synthesis and SyGuS collaborate in the synthesis process, and generate executable code that implements both reactive and data-level properties. We demonstrate the applicability of our approach over a set of real-world benchmarks [3].

#### References

1 Bernd Finkbeiner, Felix Klein, Ruzica Piskac, Mark Santolucito, *Temporal Stream Logic: Synthesis Beyond the Bools*. CAV (1) 2019: 609-629

2 Bernd Finkbeiner, Philippe Heim, Noemi Passing, *Temporal Stream Logic modulo Theories*. FoSSaCS 2022: 325-346

3 Wonhyuk Choi, Bernd Finkbeiner, Ruzica Piskac, Mark Santolucito, *Can reactive synthesis and syntax-guided synthesis be friends?*. PLDI 2022: 229-243

## 3.9 Synthesis Modulo Oracles

*Elizabeth Polgreen (University of Edinburgh, GB)*


**Joint work of** lizabeth Polgreen, Andrew Reynolds, Sanjit A. Seshia
**Main reference** Elizabeth Polgreen, Andrew Reynolds, Sanjit A. Seshia: "Satisfiability and Synthesis Modulo Oracles", in Proc. of the Verification, Model Checking, and Abstract Interpretation – 23rd International Conference, VMCAI 2022, Philadelphia, PA, USA, January 16-18, 2022, Proceedings, Lecture Notes in Computer Science, Vol. 13182, pp. 263–284, Springer, 2022.
**URL** https://doi.org//10.1007/978-3-030-94583-1_13

In classic program synthesis algorithms, such as counterexample-guided inductive synthesis (CEGIS), the algorithms alternate between a synthesis phase and an oracle (verification) phase. Many (most) synthesis algorithms use a white-box oracle based on satisfiability modulo theory (SMT) solvers to provide counterexamples. But what if a white-box oracle is either not available or not easy to work with?

In this talk, I will present a framework for solving a general class of oracle-guided synthesis problems which we term synthesis modulo oracles (SyMO). In this setting, oracles are black boxes with a query-response interface defined by the synthesis problem. This allows us to lift synthesis to domains where using an SMT solver as a verifier is not practical.

### 3.10 A primer on SYNTCOMP

*Guillermo A. Pérez (University of Antwerp, BE)*


**Joint work of** Swen Jacobs, Guillermo A. Perez
**Main reference** Swen Jacobs, Guillermo A. Pérez, Remco Abraham, Véronique Bruyère, Michaël Cadilhac, Maximilien Colange, Charly Delfosse, Tom van Dijk, Alexandre Duret-Lutz, Peter Faymonville, Bernd Finkbeiner, Ayrat Khalimov, Felix Klein, Michael Luttenberger, Klara J. Meyer, Thibaud Michaud, Adrien Pommellet, Florian Renkin, Philipp Schlehuber-Caissier, Mouhammad Sakr, Salomon Sickert, Gaëtan Staquet, Clément Tamines, Leander Tentrup, Adam Walker: "The Reactive Synthesis Competition (SYNTCOMP): 2018-2021", CoRR, Vol. abs/2206.00251, 2022.
**URL** https://doi.org//10.48550/ARXIV.2206.00251

The Reactive Synthesis Competition (SYNTCOMP) is a competition for reactive synthesis tools. The competition's goal is to collect benchmarks in a publicly available library and foster research in new tools for automatic synthesis of systems. SYNTCOMP is organized annually (since 2014) as a satellite event of CAV.

In this talk, the status of the competition (in particular the state of the benchmarks) and its evolution through the last couple of years is presented.

### 3.11 Reactive Synthesis as a Programming Language Paradigm

*Mark Santolucito (Barnard College, Columbia University – New York, US)*


**URL** https://barnard-pl-labs.github.io/CYOA-TSL

There has been an explosion of interest in the use of LLMs to generate code in recent years, complimenting a long history of formal methods-driven program synthesis. However, code generation remains a largely all-or-nothing problem – either users can take advantage of the flexibility and adaptivity of LLMs and generate code that might not be correct, or they can rely on more rigid program synthesis tools which are guaranteed to generate correct results, but are limited in their generative grammars. In this talk, we propose a strategy for the combination of these techniques – leveraging formal methods to generate structures of provable correct programs, and allowing the LLM to complete the details with more flexibility. We focus on the problem of generating reactive programs, where these types of programs consume streams of input and produce streams of output. These programs are critical across application domains, including circuit design, self-driving cars, mobile apps, and chatbots – all of which we have been able to synthesize using our synthesis procedure. We end with an outline of the challenges still facing the integration of reactive synthesis into code generation paradigms.

### 3.12 Deep Learning for Reactive Synthesis

*Frederik Schmitt (CISPA – Saarbrücken, DE)*



Neural-symbolic computing offers a broad and largely unexplored spectrum of integrating neural and algorithmic components for developing new solutions to the reactive synthesis problem. At one end of the spectrum are purely symbolic and algorithmic methods that defined the field from its very beginning. In this talk, we will move to the other end of the spectrum and discuss how far we can get by relying on pure deep learning methods to solve synthesis problems. In particular, three approaches are presented that trace the current developments in deep learning:

1. training neural networks from scratch on data derived from specification patterns,
2. fine-tuning language and code generation models,
3. evaluating large language models and few-shot prompting on instances of parameterized specifications.

We specifically focus on representing the structure of synthesis problems in neural network architectures, the bundling of both algorithmic tools and neural networks, and we hope to spark discussions about approaches that are centred on the spectrum of neural-symbolic methods.

### 3.13 The power of feedback

*Anne-Kathrin Schmuck (MPI-SWS – Kaiserslautern, DE)*



Feedback allows systems to seamlessly and instantaneously adapt their behavior to their environment and is thereby the fundamental principle of life and technology – it lets animals breathe, it stabilizes the climate, it allows airplanes to fly, and the energy grid to operate. During the last century, control technology excelled at using this power of feedback to engineer extremely stable, robust, and reliable technological systems. With the ubiquity of computing devices in modern technological systems, feedback loops become cyber-physical – the laws of physics governing technological, social or biological processes interact with (cyber) computing systems in a highly nontrivial manner, pushing towards higher and higher levels of autonomy and self-regulation. While stability, reliability and robustness remain to be of uppermost importance in these systems, a control-inspired utilization of cyber-physical feedback loops for this purpose is lacking far behind. In this talk, I will discuss how a control-inspired view on formal methods for reliable software design can enable us to utilize the power of feedback for robust and adaptable cyber-physical system design.

## 3.14 Constraint-based synthesis

*Armando Solar-Lezama (MIT – Cambridge, US)*



In this talk, I describe some recent efforts to combine functional and reactive synthesis. In the first part of the talk, I describe the Sketch program synthesis system, which allows users to write partial programs and solves for the missing details using an SMT solver. The talk focused on a new feature in Sketch that allows the user to write temporal specifications describing the behaviour of the program through its execution and showed how such constraints could be used to speed up the synthesis process and to give the user more control over the resulting program.

During the second part of the talk, I described a new effort to use a combination of reactive and functional synthesis to derive models of an environment from observations. The idea was implemented in a tool called Autumn, which focuses on pixel-world domains and is able to synthesize functional reactive programs from observations.

## 3.15 Synthesizing Pareto-optimal Interpretations for Black-box Models

*Hazem Torfah (Chalmers University of Technology – Göteborg, SE)*


**Joint work of** Hazem Torfah, Shetal Shah, Supratik Chakraborty, S Akshay, Sanjit Seshia
**Main reference** Hazem Torfah, Shetal Shah, Supratik Chakraborty, S. Akshay, Sanjit A. Seshia: "Synthesizing Pareto-Optimal Interpretations for Black-Box Models", in Proc. of the Formal Methods in Computer Aided Design, FMCAD 2021, New Haven, CT, USA, October 19-22, 2021, pp. 153–162, IEEE, 2021.
**URL** https://doi.org//10.34727/2021/ISBN.978-3-85448-046-4_24

We present a multi-objective optimization approach for synthesizing interpretations of black-box models. Existing methods for synthesizing interpretations use a single objective function and are often optimized for a single class of interpretations. In contrast, we provide a more general and multi-objective synthesis framework that allows users to choose (1) the class of syntactic templates from which an interpretation should be synthesized, and (2) quantitative measures on both the correctness and explainability of an interpretation. For a given black-box, our approach yields a set of Pareto-optimal interpretations with respect to the correctness and explainability measures. We show that the underlying multi-objective optimization problem can be solved via a reduction to quantitative constraint solving, such as weighted maximum satisfiability. To demonstrate the benefits of our approach, we have applied it to synthesize interpretations for black-box neural-network classifiers. Our experiments show that there often exists a rich and varied set of choices for interpretations that are missed by existing approaches.

# 4 Working groups

## 4.1 Quantitative Specification

*Shaull Almagor (Technion – Haifa, IL)*



Our discussion was aimed at the following question: can we find a natural specification formalism for which winning strategies in the synthesis problem are captured by well-studied quantitative computational models, such as One Counter Automata, One Counter Nets, VASS, etc.

After discussing several game types and specifications, we came up with the following concrete specification. Consider inputs $\{a, \#\}$ and outputs $\{b, @\}$, the specification is to accept only words of the form $a^n \# b^n @$. That is, the environment inputs a sequence of $a$'s, and the system should respond with a length-matching sequence of $b$'s. It is natural to model a winning strategy using a One Counter Automaton in this case. We therefore wonder if this is a particular case of a more general specification formalism. One possible candidate is "PSL with local variables", which seems to be able to capture specifications in this spirit.

## 4.2 A programmatic approach for reactive synthesis

*Nathanaël Fijalkow (CNRS – Talence, FR)*



I proposed this working group to discuss a novel approach to reactive synthesis, where instead of a finite state controller the goal is to output a controller in the form of a program in a high-level programming language.

The key questions are:

- What is the (or a?) right programming language for reactive synthesis?
- How to perform inference?
- Even model-checking of programs is not obvious

The working session gathered about a dozen participants, and the lively discussions touched upon the three key questions mentioned above. Different approaches were sketched. The matter will be investigated more thoroughly in the coming months, thanks to the interests it sparked during this working session.

## 4.3 Minimization of deterministic parity automata

*Antonio Casares (University of Bordeaux, FR)*



In this working group, we discussed the complexity of the following decision problem: Input: A deterministic transition-based parity automaton $\mathcal{A}$ and an integer $k$. Question: Is there a deterministic transition-based parity automaton of size at most $k$ equivalent to $\mathcal{A}$?

We considered the minimization procedure for good-for-games coBüchi automata by Abu Radi and Kupferman [1], and studied what are the kind of combinatorial problems that arise when trying to determinize these automata by adding a minimal number of states.

**References**

1 Bader Abu Radi, Orna Kupferman. *Minimization and Canonization of GFG Transition-Based Automata*. Log. Methods Comput. Sci., 2022

## 4.4 Minimization of deterministic (co)Büchi automata

*Rémi Morvan (University of Bordeaux, FR)*



This working group is a follow-up of Antonio Casares' session "Minimization of deterministic parity automata". We studied the following functional problem:

**input:** A deterministic transition-based (co)Büchi automaton $\mathcal{A}$.

**output:** A deterministic transition-based (co)Büchi automaton which is equivalent to $\mathcal{A}$, and state-minimal.

Contrary to minimization of parity automata, we believe this problem to be polynomial-time computable. We mostly focused on a congruence-based approach.

## 4.5 Positionality and memory

*Pierre Ohlmann (University of Warsaw, PL)*



We worked on understanding memory requirements in games with topologically open winning conditions. The question can be phrased as follows. Let $\Sigma$ be an alphabet:

**input:** $L$ a language of $\Sigma^*$, say regular;

**output:** minimal $k$ such that on any game graph with objective $L \subseteq \Sigma^\omega$, if Eve wins then she wins with a $k$-states memory

We discussed a few examples and motivations and talked about the case where $L$ is a singleton which has a simple solution. We did not make substantial progress on the general case.

## 4.6 Graph neural networks and reactive synthesis

*Guillermo A. Pérez (University of Antwerp, BE)*



**Joint work of** Guillermo A. Perez, Isseinie Calviac

During this working session, we discussed the possibility of using graph neural networks (GNNs) to reduce the state space of automata used for synthesis from LTL specifications. The state of the art (in current LTL-synthesis solvers) concerns the syntactic recognition of subformulas that are treated with special transformations which may lead to minimal subautomata. Finally, the product of said automata is taken to obtain a final automaton. To further reduce the size of the final automaton, Spot /ltlsynt implements an approximation of an isomorphism check (rather, it resembles graded bisimulation). This option seems to be disabled by default as it takes too long. An alternative concerns guessing a bisimulation relation and checking (in logspace, using one step of the classical partition refinement algorithm for bisimulation) that it is indeed a bisimulation relation. This begs the question of whether machine learning can help us make such a guess.

On the GNN side, it was mentioned that they are a very hot topic in AI and that the main focus of current research concerns the proposal of new architectures to improve their "expressiveness". GNNs can be seen as a "recolouring" function applied to a coloured graph. It is known that this recolouring function cannot colour nodes differently that the Weisfeiler-Leman (WL) algorithm would deem as equivalent. Conversely, for every coloured graph, there are weight matrices and bias vectors such that GNNs implement the WL algorithm. Guillermo A. Perez mentioned an unpublished result: GNNs can also implement the (coarsest) bisimulation relation.

Open questions:

1. Can other relations of interest be implemented using different GNN architectures?
2. To leverage GNNs in graphs that come from verification applications, one needs to obtain useful and meaningful features for the vertices. How can these be obtained easily or even automatically?
3. For large graphs, recent works and libraries suggest sampling a number of neighbours of each node. Will this render GNNs useless for verification methods in large graphs?

### References

1 Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, Martin Grohe: Weisfeiler and Leman Go Neural: Higher-Order Graph Neural Networks. AAAI 2019: 4602-4609

2 Floris Geerts, Filip Mazowiecki, Guillermo A. Pérez: Let's Agree to Degree: Comparing Graph Convolutional Networks in the Message-Passing Framework. ICML 2021: 3640-3649

3 Tobias Hecking, Swathi Muthukrishnan, Alexander Weinert. Predicting Winning Regions in Parity Games via Graph Neural Networks. Deep Learning-aided Verification @ CAV 2023.

## 4.7 SYNTCOMP benchmarking

*Guillermo A. Pérez (University of Antwerp, BE)*


**Joint work of** Swen Jacobs, Guillermo A. Perez, Philipp Schlehuber-Caissier

Benchmark sets for existing and upcoming tracks of the reactive synthesis competition (SYNTCOMP) were discussed. The main points that were touched during the discussion can be summarized as follows.

1. PDDL is a well-established family of languages in the planning community for describing tasks. In the short term, we can try to translate (non-deterministic) PDDL specifications to TLSF. In the long term, PDDL itself may become a reasonable format for some tracks, and we may further extend it with goals expressed in LTL over finite words.

2. While reports of SYNTCOMP usually include graphs covering all benchmarks, it may be of more importance to highlight how well tools scale per parametric family of benchmarks as the parameter values increase. This is already present in the report for the 2018-2021 editions of the competition. A proposal is to include such graphs in the results of every forthcoming edition of the competition.
3. PSL is a well-established extension of LTL that seems to be used in industry. We will survey the literature on PSL case studies and perhaps even approach IBM, Synopsys, and other companies to ask whether they have such specifications so that we can enrich the set of benchmarks.
4. Regarding the output format of synthesis tools: The current one is quite succinct and it matches the input of model checking tools. Namely, SYNTCOMP requires output in AIGER format. However, semi-explicit representations such as HOA for Mealy/Moore machines may be easier to visualize. Hence, for future editions of the competition, an "explainable badge" will be awarded to tools that produce such semi- or fully explicit versions of their output (as an additional option, not as the official output for the competition).
5. Finally, we noted that the parity game track is currently biased as it measures the time for parsing and minimization of the parity game together with the time it takes to actually solve the game. Instead, a proposal is to split this into preprocessing time and solving time. In the short term, tools will have to output a message and timestamp stating when preprocessing is finished so that we can split the two processes. In the long term, we can consider more succinct formats of a parity game to have parsing become faster (after using, perhaps, the BDD version of hoa-tools) and to be able to succinctly encode even larger games.

## 4.8 IPASIR-UP: User Propagators for CDCL

*Andre Schidler (TU Wien, AT)*


**URL** https://simons.berkeley.edu/talks/katalin-fazekas-tu-wien-2023-04-17

This talk is a teaser talk about a new SAT solver API – IPASIR-UP – that provides interactions with the solver during the solving process. Hence, instead of calling the solver incrementally, it is possible to interact with the solver during the solving process, i.e., whenever the solver makes a decision, propagates a variable, finds a conflict, finds a model, etc.

IPASIR-UP allows, among other things, implementing CEGAR approaches or theories quickly and cleanly. So far, this API is available in Cadical.

## 4.9 Reactive Synthesis Beyond the Bools

*César Sánchez (IMDEA Software Institute – Madrid, ES)*



There has been a growing interest in the last few years in increasing the expressivity of reactive synthesis from propositional LTL into richer languages that can handle data. One impact-full line of research includes temporal stream logic (TSL) that has managed to synthesize sophisticated controllers. Another example was Raina Dimitrova's work presented in this seminar. However, most attempts to increase the expressivity quickly render the realizability decision problem undecidable.

In the first part of this working session we discussed the recent work on realizability of "LTL modulo theory" that shows decidability of an extension of LTL where the propositions are replaced by literals from a first order theories. The Boolean abstraction method generates an equi-realizable propositional LTL specification as long as the first-order theory enjoys a decidable exists-forall (validity) decision problem. Moreover, the resulting specification remains in the same temporal class and is amenable of Boolean reactive synthesis. Then we discussed how the resulting (Boolean) controllers obtained using existing synthesis tools can be extended to "theory controllers", thus obtaining a full LTL modulo theory synthesis algorithm.

However, current LTL modulo theory does not allow to transfer data across time (which is the main reason for undecidability of richer formalisms like TSL). We briefly discussed possibilities for enriching LTL modulo theories with controlled data transferred, particularly based on known decidability results from register automata. We concluded that a first promising approach should be based on specific characteristics of each theory and not (as for LTL modulo theories described above) agnostic to the theory in question.

## 4.10 Reactive synthesis of Linear Temporal Logic on finite traces

*Shufang Zhu (University of Oxford, GB)*



In this working group, we discussed the following problem in reactive synthesis of Linear Temporal Logic on finite traces (LTLf):

Consider an autonomous system immersing in an environment, where there are multiple abstraction levels of how the environment behaves, depending on how much environment dynamics to be concerned:

- $\text{Env}_1$ (any possible env behaviours),
- $\text{Env}_2$,
- …
- $\text{Env}_m$ (most restricted env behaviours).

The system is given multiple tasks, ranging from most difficult to easiest:

- $\text{Task}_1$ (most difficult),
- $\text{Task}_2$,
- …
- $\text{Task}_n$ (easiest).

Question: How to achieve a good balance to achieve a more difficult task considering a more flexible environment?

We discussed different ways of dealing with this problem. A good direction might be looking at robust LTL [1]. Another interesting direction is MaxSAT or Weighted MaxSAT.

### References

1 Paulo Tabuada, Daniel Neider. *Robust Linear Temporal Logic*

## Participants


- Shaull Almagor
Technion – Haifa, IL
- Guy Avni
University of Haifa, IL
- Mrudula Balachander
Free University of Brussels, BE
- Véronique Bruyère
University of Mons, BE
- Michaël Cadilhac
DePaul University – Chicago, US
- Antonio Casares
University of Bordeaux, FR
- Rayna Dimitrova
CISPA – Saarbrücken, DE
- Alexandre Duret-Lutz
EPITA – Le Kremlin Bicêtre, FR
- Rüdiger Ehlers
TU Clausthal, DE
- Nathanaël Fijalkow
CNRS – Talence, FR
- Emmanuel Filiot
UL – Brussels, BE
- Bernd Finkbeiner
CISPA – Saarbrücken, DE
- Dana Fisman
Ben Gurion University – Beer Sheva, IL
- Hadar Frenkel
CISPA – Saarbrücken, DE
- Swen Jacobs
CISPA – Saarbrücken, DE
- Ayrat Khalimov
TU Clausthal, DE
- Bakh Khoussainov
Univ. of Electronic Science & Technology – Chengdu, CN
- Rupak Majumdar
MPI-SWS – Kaiserslautern, DE
- Théo Matricon
University of Bordeaux, FR
- Niklas Metzger
CISPA – Saarbrücken, DE
- Rémi Morvan
University of Bordeaux, FR
- Anca Muscholl
University of Bordeaux, FR
- Pierre Ohlmann
University of Warsaw, PL
- Guillermo A. Pérez
University of Antwerp, BE
- Ruzica Piskac
Yale University – New Haven, US
- Elizabeth Polgreen
University of Edinburgh, GB
- Mickael Randour
F.R.S.-FNRS & UMONS – Université de Mons, BE
- César Sánchez
IMDEA Software Institute – Madrid, ES
- Mark Santolucito
Barnard College, Columbia University – New York, US
- Andre Schidler
TU Wien, AT
- Frederik Schmitt
CISPA – Saarbrücken, DE
- Anne-Kathrin Schmuck
MPI-SWS – Kaiserslautern, DE
- Martina Seidl
Johannes Kepler Universität Linz, AT
- Armando Solar-Lezama
MIT – Cambridge, US
- Hazem Torfah
Chalmers University of Technology – Göteborg, SE
- Tom van Dijk
University of Twente – Enschede, NL
- Shufang Zhu
University of Oxford, GB
- Martin Zimmermann
Aalborg University, DK