DAGSTUHL
REPORTS

Volume 6, Issue 2, February 2016

Dagstuhl Reports, Vol. 6, Issue 2 ISSN 2192-5283

ISSN 2192-5283

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

Publication date
September, 2016

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

Digital Object Identifier: 10.4230/DagRep.6.2.i

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

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

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

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

Report from Dagstuhl Seminar 16061

Data-Driven Storytelling

Edited by

Sheelagh Carpendale[1], Nicholas Diakopoulos[2], Nathalie Henry Riche[3], and Christophe Hurter[4]

1 **University of Calgary, Calgary, CA,** sheelagh@ucalgary.ca
2 **University of Maryland – College Park, US,** nicholas.diakopoulos@gmail.com
3 **Microsoft Research – Redmond, US,** nathalie.henry@microsoft.com
4 **ENAC – Toulouse, FR,** christophe.hurter@enac.fr

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16061 "Data-Driven Storytelling". Close to forty researchers and practitioners descended on Schloss Dagstuhl to forge an interdisciplinary agenda on the topic of data-driven storytelling using visualization in early February, 2016. With burgeoning research interest in understanding what makes visualization effective for communication, and with practitioners pushing the envelope of the craft of visual communication, the meeting put different modes of thinking between computer science researchers and data visualization practitioners in close proximity for a week.

Seminar February 7–12, 2016 – http://www.dagstuhl.de/16061
1998 ACM Subject Classification H.5 Information Interfaces and Presentation, H.5.1 Multimedia Information Systems
Keywords and phrases data journalism, information visualization, personal visualization, storytelling, visual literacy
Digital Object Identifier 10.4230/DagRep.6.2.1
Edited in cooperation with Alice Thudt

1 Summary

Sheelagh Carpendale
Nicholas Diakopoulos
Nathalie Henry Riche
Christophe Hurter

Data visualization is the "use of computer-supported, interactive, visual representations of data to amplify cognition" [5]. Visualization can play a crucial role for exploring data and for communicating information as "a picture is worth a thousand words". Early research in this field focused on producing static images and quantifying the perception of different visual encodings [6] in these visual representations. The vast majority of research since then focused on designing and implementing novel interfaces and interactive techniques to enable data exploration. Major advances in visual analytics and big data initiatives concentrated on integrating machine learning and analysis methods with visual representations to enable powerful exploratory analysis and data mining [10]. As interactive visualizations play an increasing role in data analysis scenarios, they also started to appear as a powerful vector for communicating information. Stories supported by facts extracted from data analysis proliferate in many different forms from animated infographics and videos [2] to interactive online visualizations on news media outlets. We argue that it is now time for the visualization research community to understand how these powerful interactive visualizations play a role in communicating information. We define this line of research as data-driven storytelling.

The popularity of javascript web technology and the availability of the D3 toolkit [3] enabled a wider range of people to create data visualizations. Being able to easily share interactive data visualizations on the web also increased the democratization of interactive visualizations. Coupled with the emphasis on data science, these advances raise new practices such as data journalism. Data journalists gather and explore available datasets to extract relevant insights, often conveying their stories via interactive data visualizations [1, 9]. The popularity of data-driven stories on New York Times especially, revealed the potential of interactive visualizations as a powerful communication tool [7].

Central to our vision of the convening was that the vast majority of research on data visualization to date has focused on designing and implementing novel interfaces and interactive techniques to enable data exploration. Major advances in visual analytics and big data initiatives have concentrated on integrating machine learning and analysis methods with visual representations to enable powerful exploratory analysis and data mining. But just as interactive visualization plays an important role in data analysis scenarios it is also becoming increasingly important in structuring the communication and conveyance of insights and stories in a compelling format. Visual data-driven stories have proliferated in many different forms, from talks [8], to animated infographics and videos [1, 9, 7], to interactive online visualizations.

Data-driven storytelling is also compelling for a wide range of applications. In enterprise scenarios, the output of data analysis (often reports and slide-based presentations) has to be conveyed to decision makers. In scientific research, interactive visualizations are increasingly used to convey data-driven discoveries to peers or used to communicate complex findings to a broader audience. In education scenarios, interactive visualizations are used by teachers to explain mathematical concepts or to illustrate biological or physical mechanisms. Many questions arise as interactive visualizations are used beyond data exploration by experts, for communication purposes to a broader audience. Research on understanding of static images in cognitive psychology and perception must be extended to encompass more advanced techniques (videos and interactive applications). Visualization literacy, defined as the ability to extract, interpret, and make meaning from information presented in the form of an (interactive) data visualization is also a crucial component for data-driven storytelling research. Assessing the visualization literacy of an audience and developing techniques to better teach how to decode interactive visualizations has started to attract the attention of our research community [4] However a plethora of research remains to be done. For example, research on how visualizations can lie [11] or at least how they may introduce bias in the reader's mind has focused on static visual representations but has not yet been extended to other medium. Similarly it is crucial for advancing researches in visualization to assess the role data-driven storytelling can play in easing the comprehension of a messages or in increasing their memorability.

The visualization research community needs to reflect on data-driven storytelling and to develop a research agenda to investigate how advanced data-driven stories are understood by the audience, identify factors that makes them compelling as well as factors that can introduce bias in their perception. By learning from master storytellers from other fields (journalism, design, art and education) strategies to craft successful stories, our community will be able to reflect on these questions and eventually build novel consuming tools that

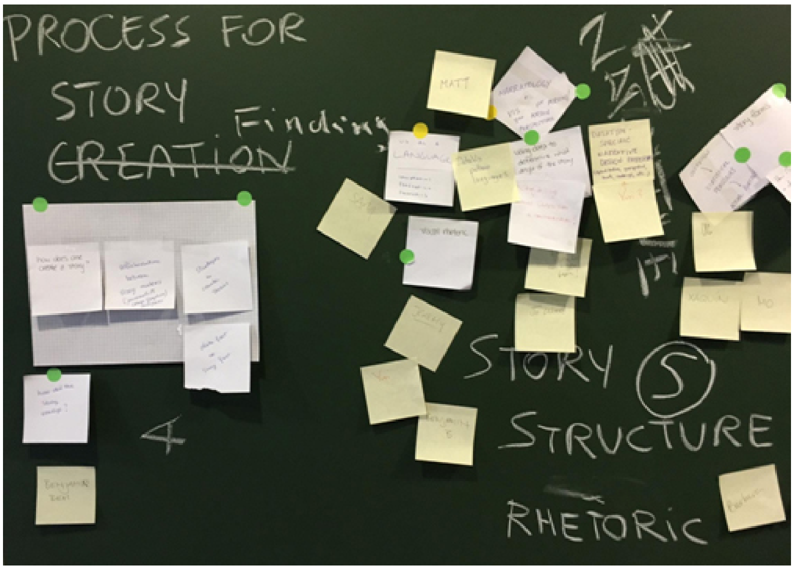

Figure 1 Converging on topical groups from hundreds of individual ideas.

engage a broad audience while minimizing perception bias, as well as build novel authoring tools to craft high quality data-driven stories.

One domain where there has been extensive and practical progress on the question of data-driven storytelling is data journalism. News sites like FiveThirtyEight or the New York Times' The Upshot have seen a recent surge of attention and interest as a means of communicating data-driven news to the public. By carefully structuring the information and integrating explanation to guide the consumer, journalists help lead users toward a valid interpretation of the underlying data. Because of the rapid and practical progress of data-driven storytelling in the domain of journalism, our seminar sought to put some of the top practitioners from that field together with computer science researchers to discuss the challenges and opportunities of data-driven communication.

The Dagstuhl seminar was structured to leverage the interdisciplinarity of the attendees by first tapping into a divergent design thinking process meant to enumerate the range of issues that are relevant to data-driven stories. Hundreds of index cards and sticky notes were sacrificed as participants generated ideas (see Figure 1).

We then clustered these ideas to arrived at a set of key themes, including:

- Techniques and Design Choices for Storytelling
- Exploration and Explanation
- From Analysis to Communication
- Audience
- Evaluation
- Devices and Gadgets
- Ethics

Groups of participants formed around common interests and each of these major themes were then the focus of discussion. Each work group was geared towards developing an outline and plan to produce a written chapter for a forthcoming edited book on the topic of data-driven storytelling. Some groups met for a day or two and then reformed around other topics, whereas other groups spent the entire week going deep in exploring a single topic. And as if the daytime activities weren't enough, additional evening breakout groups formed around additional topics of interest like Education in Data Visualization, Urban Visualization, and the Technology Stack for data-driven stories.

In-between the intense, small group sessions the entire group came together daily for five-minute lightning talks on a wide array of relevant topics. These stimulating talks primed the group for approaching data-driven storytelling from different perspectives and were an entertaining and informative way to share creative ideas or results in small and easily digestible nuggets. Among the more than 25 lightning talks, topics ranged from storytelling with timelines, to mobile visualization, the use of data comics, visual literacy, affect and color, data-story design workflows, and even the visualization of data through cuisine.

Outcomes

Our initial goal of the seminar was to have groups work intensively on their chosen topic(s) so that an outline and workplan could be developed to write a contributing chapter to a book on data-driven storytelling. The book is underway and will have contributions on each of the main themes outlined above, as well as an introductory chapter by the editors / organizers of the Dagstuhl seminar. Moreover, our creative contributors at the seminar produced other outputs as well: curated lists of example data driven stories, as well as of storytelling techniques were created and will be published online, and a blog has pulled together some of the formative impressions of participants (https://medium.com/data-driven-storytelling).

Below we briefly summarize the expected contents of each of the chapters that will form the book.

Techniques and Design Choices for Storytelling

This chapter will discuss techniques and design choices for visual storytelling grounded in a survey of over 60 examples collected from various online news sources and from award-winning visualization and infographic design work. These design choices represent a middle ground between low-level visualization and interaction techniques and high-level narrative devices or structures. The chapter will define several classes of design choices: embellishment, explanation, exploration, navigation, story presentation, emphasis, focus, and annotation. Examples from the survey for each class of design choices will be provided. Finally, several case studies of examples from the survey that make use of multiple design choices will be developed.

Exploration and Explanation in Data-Driven Stories

This chapter will explore the differences between and integration of exploration and explanation in visual data-driven storytelling. Exploratory visualizations allow for a lot of freedom which can include changing the visual representation, the focus of what is being shown and the sequence in which the data is viewed. They allow readers to find their own stories in the data. Explanatory stories include a focused message which is usually more narrow and guides the reader often in a linear way. Advantages and disadvantages of exploration and explanation as well as dimensions that help to describe and classify data-driven stories will be developed. The space is described by identifying freedom, guidance regarding representation, focus and sequence as well as interpretation as important dimensions of data-driven storytelling and existing systems are characterized along these dimensions. Recommendations will be developed for how to integrate both aspects of exploration and explanation in data-driven stories.

From Analysis to Communication: Supporting the Lifecycle of a Story

This chapter will explore how tools can better support the authoring of rich and custom data stories with natural / seamless workflows. The aim is to understand the roles and limitations of analysis / authoring tools within current workflow practices and use these insights to suggest opportunities for future research and design. First, the chapter will report a summary of interviews with practitioners at the Dagstuhl seminar; these interviews aim to understand current workflow practices for analysis and authoring, the tools used to support those practices, and pain points in those processes. Then the chapter will reflect on design implications that may improve tool support for the authoring process as well as research opportunities related to such tool support. A strong theme is the interplay between analytical and communicative phases during both creation and consumption of data-driven stories.

The Audience for Data-Driven Stories

Creators of data-driven visual stories want to be as effective as possible in communicating their message. By carefully considering the needs of their audience, content creators can help their readers better understand their content. This chapter will describe four separate characteristics of audience that creators should consider: expertise and familiarity with the topic, the medium, data, and data visualization; expectations about how and what the story will deliver; how the reader uses the interface such as reading, scrolling, or other interactivity; and demographic characteristics of the audience such as age, gender, education, and location. This chapter will discuss how these audience goals match the goals of the creator, be it to inform, persuade, educate, or entertain. Then it will discuss certain risks creators should recognize, such as confusing or offending the reader, or using unfamiliar jargon or technological interfaces. Case studies from a variety of fields including research, media, and government organizations will be presented.

Evaluating Data-Driven Storytelling

The study of data-driven storytelling requires specific guidelines, metrics, and methodologies reflecting their different complex aspects. Evaluation is not only essential for researchers to learn about the quality of data-driven storytelling but also for editorial rooms in media and enterprises to justify the required resources the gathering, analyzing and presentation of data. A framework will be presented that takes the different perspectives of author, audience and publisher and their correspondent criteria into account. Furthermore it connects them with the methods and metrics to provide a roadmap for what and how to measure if these resulting data-driven stories met the goals. In addition, the chapter will explore and define the constraints which might limit the metrics and methods available making it difficult to reach the goals.

Devices and Gadgets for Data Storytelling

This chapter will discuss the role of different hardware devices and media in visual data driven storytelling. The different form factors offer different affordances for data storytelling affecting their suitability to the different data storytelling settings. For example, wall displays are well suited to synchronous co-located presentation, while watches and virtual reality headsets work better for personal consumption of pre-authored data stories.

Ethics in Data-Driven Visual Storytelling

Is the sample representative, have we thought of the bias of whoever collected or aggregated the data, can we extract a certain conclusion from the dataset, is it implying something the data doesn't cover, does the visual device, or the interaction, or the animation affect the interpretation that the audience can have of the story? Those are questions that anyone that has produced or edited a data-driven visual story has, or at least should have, been confronted with. After introducing the space, and the reasons and implications of ethics in this space, this chapter will look at the risks, caveats, and considerations at every step of the process, from the collection/acquisition of the data, to the analysis, presentation, and publication. Each point will be supported by an example of a successful or flawed ethical consideration.

Conclusion

The main objective of this Dagstuhl seminar was to develop an interdisciplinary research agenda around data-driven storytelling as we seek to develop generalizable findings and tools to support the use of visualization in communicating information. Productive group work converged to delineate several research opportunities moving forward:

- The need for interfaces that enable the fluid movement between exploratory and communicative visualization so that storytelling workflow is seamless and powerful.
- The need to develop typologies of visual storytelling techniques and structures used in practice so that opportunities for supporting these techniques can be sought through computing approaches.
- The need to develop evaluation frameworks that can assess storytelling techniques and tools both scientifically and critically.
- The need for design frameworks that can guide the structure of visual information for experiences across different output devices, both existing and future.
- The need to understand the audience and their role in co-constructing meaning with the author of a data-driven story.
- The need for ethical frameworks that should guide tool development for visual data-driven communication.

These opportunities were productively enumerated at the Dagstuhl seminar and are in the process of being written up as chapters in our book on data-driven storytelling.

References

1 All the medalists: Men's 100-meter freestyle: Racing against history. Web, 2012. http://www.nytimes.com/interactive/2012/08/01/sports/olympics/racing-against-history.html?_r=1&.

2 Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Christophe Hurter, and Pourang Irani. Understanding data videos: Looking at narrative visualization through the cinematography lens. In *Proceedings of CHI: SIGCHI Conference on Human Factors in Computing System*, pages 1459–1468, New York, NY, USA, 2015. ACM.

3 Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, December 2011.

4 Jeremy Boy, Sara Johansson Fernstad, Martin Turner, Simon Walton, David Ebert, Jean-Daniel Fekete, Andy Kirk, and Mario Romeo. Eurovis 2014 workshop: Towards visualization literacy. visualization literacy workshop. EuroVis 2014, 2014. https://www.kth.se/profile/178785/page/eurovis-2014-workshop-towards-visualiza/.

5 Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA., 1999.

6 William C. Cleveland and Marylyn E. McGill. *Dynamic Graphics for Statistics (1st ed.)*. CRC Press, Inc., Boca Raton, FL, USA., 1999.

7 Instance of data storytelling to explain the subprime mortgage crisis in the usa. Web, 2014. http://www.bloomberg.com/dataview/2014-02-25/bubble-to-bust-to-recovery.html.

8 Hans Rosling. The best stats you've ever seen. TED Talks, 2006. https://youtu.be/hVimVzgtD6w.

9 The wealth report 2013: Examining high net worth individuals from around the world. Web, 2013. http://www.elitehavenssales.com/news/47/The-Wealth-Report-2013-Examining-High-Net-Worth-Individuals-from-around-the-world.

10 J. J. Thomas and K. A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE CS Press, USA., 2005.

11 Edward Tufte. *The Visual Display of Quantitative Information, Second Edition*. Graphics Press, USA., 1991.

2 Table of Contents

Working groups

3 Overview of Talks

3.1 Data Clips

Fereshteh Amini (University of Manitoba – Winnipeg, CA)

Data videos and animated infographics are becoming an increasingly popular medium for conveying digital stories with data. However, data video creation is challenging as it requires one to hone a unique and combined set of skills in areas such as storyboard design, data visualization, and motion graphics. In this project, we introduce DataClips, an authoring tool designed to help craft animated infographics. We designed DataClips by first undergoing a qualitative examination of existing data videos, authored by reputable sources. This provided us with a rich set of the most important and common visual properties that define a data-driven clip, or what can be considered as being as an atomic unit of a data video.

3.2 Telling Stories about Dynamic Networks with Graph Comics

Benjamin Bach (Microsoft Research – Inria Joint Centre, FR)

Comics are a well known medium to communicate stories by unfolding temporal events in visual space. We explore the design space of comics to present insights into changes in networks.and other visual languages, such as weather charts and assembly instruction, to present insights into changes in networks. However, while comics evolved around stories in "the real world", depicting changes in data (here, dynamic networks) poses several challenges: how to identify data elements across panels? how to switch between overview and detail, how to convey importance of data elements, and so forth. To understand the potential of comics as a storytelling medium, we first created a variety of comics during a 3 month structured design process, involving domain experts from public education and neuroscience. This process led to the definition of 8 design factors for creating graph comics and propose design solutions for each. Results from a qualitative study suggest that a general audience is quickly able understand complex temporal changes through graph comics, provided with minimal textual annotations and no training.

3.3 Affect in vis: Colour and animation

Lyn Bartram (Simon Fraser University – Surrey, CA)

The creation of affect – feelings, experiences, or impressions – has a central role in creating immersive and engaging stories. But affect is also important in the context of how an evoked experience or feeling amplifies, augments or moderates the cognitive interpretation of a visualization: it can enhance interpretation, promote engagement and appropriately "frame" the information. We are investigating how different properties of color and animation may contribute to affective interpretation of visualizations with the objective of increasing the expressive capacity of visual representation.

3.4 The Works: All the way from data collection to visualizations in Subspotting

Dominikus Baur (Andechs, DE)

Subspotting mapped the cell phone reception on the New York City subway. While officially there is none available, stray signals from aboveground let people access the network at certain places on certain lines. In this project, Daniel Goddemeyer and Dominikus Baur collected and visualized the reception quality on the full subway and turned it into an iPhone app, two posters and an interactive website.

3.5 Thoughts on Visualization literacy

Jeremy Boy (NYU Polytechnic School of Engineering, US)

In this lightning talk I present my thoughts on visualization literacy. I introduce the different building blocks of a common visualization to illustrate what needs to be understood in order to comprehend the visualization. I provide my definition of visualization literacy, and propose a series of perspectives for ongoing and future research. I discuss other takes on literacy, specifically that it is more than simply reading or understanding: it also about being able to write or author visualizations. Finally, I suggest that there are different levels to the concept of visualization literacy, spanning from the low level immediate comprehension of graphics to the higher level comprehension of expressive forms like rhetoric, persuasion, deception or poetry.

3.6 Storytelling with timeline data

Matthew Brehmer (University of British Columbia – Vancouver, CA)

This is a talk about storytelling with timeline or event sequence data. The dominant form of timeline presentation is a linear, chronological encoding pioneered by Joseph Priestley in 1765. But this is certainly not the only form of timeline encoding. Based on a survey of over 250 timeline infographics and timeline visualization tools, we proposed a design space for timelines with three dimensions: representation, scale, and layout. This design space allows a designer or storyteller to present a different narrative points related to different aspects of timeline data, as some combinations of dimensions from our design space are better suited to communicate different narrative points. We developed an environment for creating timelines that combine these dimensions. But what if you wanted to tell a story that involves more than one type of narrative point? For this reason, we also explored the possibilities for animating timelines, to allow authors to create stories that accommodate different narrative points without loss of context. I present an example story produced using our environment.

3.7 Demo of SandDance using storytelling with unit visualizations

Steven M. Drucker (Microsoft Research – Redmond, US)

SandDance is a visualization system that blends between exploration and presentation. It uses the notion of representing every data element on the screen in a series of visualizations with animated transitions that help both the presenter and the audience maintain their context as different representations highlight different aspects in the data.

3.8 Network Visualisation: a “higher-order” visual analytics for data exploration and communication

Tim Dwyer (Monash University – Caulfield, AU)

Visualisation of quantitative data using basic chart graphics is well understood in its applicability to the analysis of low-dimensional data and time-series. It is clear that people are able to obtain insights (such as patterns, trends, clustering, outliers and so on) through visual inspection of a very simple mapping of such data to a picture. However, machine learning techniques are also very good at identifying such features in simple data. If the ultimate aim of such data analysis is to make connections and build knowledge, then perhaps we should be trying harder to leapfrog the simple mappings of data to chart graphics and visualise the knowledge directly to support a higher-order form of reasoning about data. In this discussion we posit that network visualisation can be used as such a “higher-order” analysis tool. We examine some use-cases where this is already being attempted and opportunities for the visualisation community to further explore this space.

3.9 Towards cultural sensitivity/sensibility: Visualizing texture and structure of cultural collections

Marian Dörk (Fachhochschule Potsdam, DE)

The research project “Visualizing Cultural Collections” (VIKUS – Visualisierung kultureller Sammlungen) investigates graphical user interfaces and the potentials inherent to visual exploration of digitized cultural collections. Within the scope of the project, researchers from various fields such as information visualization, interface design, and cultural sciences, will collaborate to develop and evaluate graphical interfaces for interactive exploration of cultural objects. Recently, there has been a particular research interest in representations that show it all, while still revealing the individual pieces. Examples include SandDance (Steven Drucker et al.) and Image plots (Lev Manovich et al.). I have demonstrated the prototype visualization “PAST VISIONS penned by Frederick William IV”, which shows the collection of historical drawings by Frederick William IV alongside a thematic and temporal arrangement. His drawings are part of the collection of prints and drawings of the Prussian Palaces and Gardens Foundation Berlin-Brandenburg (SPSG) and have previously been published in an online inventory catalogue. Cultural collections poses some unique questions for narrative in visualization: How to pursue the risk of getting lost as an opportunity for open engagement with our cultural heritage? Can we create the opportunity for various readings? In which ways can we treat visualization as both method and object of interpretation?

3.10 10x10 stories

Kennedy Elliott (The Washington Post, US)

My unique perspective on data-driven storytelling comes from my own field of news graphics and news data journalism. The key to our work is simply telling stories that connect with readers – intellectually, emotionally and/or visually. For every project, we consider whether the reader would likely understand the key editorial components of the visual story, if the technical design leads the reader through the story effectively and finally if the technical development is intuitive, bug-free and suits the reader on any particular device. Our data-driven stories involve data types that range from conventional to more abstract: from continuous values to pictures, videos and satellite imagery.

3.11 Thoughts on breaking up the linearity and passivity of video storytelling

Kennedy Elliott (The Washington Post, US)

News organizations and content producers have yet to come up with effective ways to integrate video into textual, data and visual stories. Video is passive and poses and interruption to an otherwise active activity of reading or interacting with a story online. Readers have no control over the speed of progression of the video and surrender to the pace the video storyteller intended. Audio can also be a deterrent in choosing to watch a video. I will present a number of possible solutions to these long-standing problems of video and how they integrate into stories.

3.12 An area of conflicts: The antagonists and promoting factors of dataviz projects in newsrooms

Christina Elmer (Spiegel Online GmbH – Hamburg, DE)

Newsrooms are special ecosystems with their own sometimes strong mechanisms and values system. We as dataviz practitioners should be aware of the forces arising from that. We should use promoting factors like beauty, virality, uniqueness, transparency, personalization and impression of depth – as arguments to gather and to justify the resources and manpower we need for those projects. On the other hand, there are some antagonists that can make it difficult to produce such pieces within newsrooms. Many news topics have to be finished within deadlines, sometimes hours after the investigation starts. The cost pressure in many news outlets limits the resources for dataviz projects and there also is still a quite remarkable gap between technology and journalism to be overcome. The rising mobile usage of news websites influences the possible size and granularity of the visual contents. Thus, both the usual priorities and workflows have to be considered and adjusted to data-driven projects – to enable success within the system and a sustainable change towards more dataviz in the news.

3.13 Practicalities of visual storytelling on the smaller, connected, portable canvas

Xaquin Gonzalez Veira (The Guardian Visuals – London, GB)

In the last year (2015) the Guardian has effectively transition to a mobile first workflow for its interactive storytelling and data visualizations. In that time, we have notice a few patterns and collected a few considerations. There are obvious limitations of the size the canvas itself and landscape v. portrait orientations, but also optimizations on load, limits for autoplaying videos, different in-app browsers, different time commitments, advantages of gestures on touch screens (like in the 'Build your own coalition' interactive) ... We have also learned that some projects may require a completely different expression of the content. In the 'Mekong River' interactive documentary, the desktop experience is simply a long scrolling page and uses autoplay videos on scroll; the mobile expression, which was developed and conceptualized first, is a set of cards that give you a curated story when swiping vertically, and at certain points, the options of finding out more about that section when swiping horizontally.

3.14 Visualizing the news at The Guardian

Xaquin Gonzalez Veira (The Guardian Visuals – London, GB)

In 2014, The Guardian merged the graphics and interactive teams, motion graphics specialists, interactive documentarians and the picture desk – with a renewed emphasis on digital – under the umbrella of Visuals. The idea is to push for a holistic story editing: each element graphics, text, video, pictures ... works depending on how it plays in the context of the whole story. It also means a renewed effort on interactive data visualizations, and the size of the department gives it the opportunity to be proactive in finding newsworthy stories. The interactives 'How China's economic slowdown could weigh on the rest of the world' and 'Unaffordable country: Where can you afford to buy a house?', both in collaboration with the also newly created Data team, are an example of our data-driven visual storytelling philosophy, and among other tenets: an emphasis on new angles for stories and a search for affect in the connection between the visuals and the text.

3.15 Using Data-Driven Storytelling to Make Data and Science Relatable

Jessica Hullman (University of Washington – Seattle, US)

As data and scientific results are increasingly prevalent in public spheres, we need tools to help demystify difficult aspects of data and science for interested non-domain experts. I'll describe tools and mechanisms I'm developing that formalize strategies that the best human narrators rely on explain complex concepts. Our work in scientific text simplification learns mappings between complex scientific terms and simpler equivalents from both scientific journals and science news articles. A user of our interactive authoring and reading applications can access simplifications on demand as they read or write about science. Our tools for supporting measurement comprehension take an input measurement that is unfamiliar to a user and output analogies that relate that measurement to familiar objects and landmarks (300 acres is about the area of a well-known park in your town). Our work in visualizing uncertainty shows that presenting uncertainty as an animated set of draws provides a more directly experienceable, countable, and ultimately more understandable experience of probability than static plots of distributions.

3.16 Visualization assembly process as a story?

Samuel Huron (ENST – Paris, FR)

In the last few years I saw more and more data-driven video on YouTube where people are assembling visualization physically or visually. In these videos, people are using movable independent visual elements map to data unit (I named tokens). These elements could be pictograms, pixels (like in SandDance), stones (like Hans Rosling shortest ted talk), lego bricks, bubble gums and others. In most of these videos the authors support a data-driven narrative, by assembling seamlessly these elements over time to construct more complex representations. During my lightning talk, I explored the following question: How and why the temporal assembly process of such constructed visualization can be a visual story process? To discuss this question, I compared commonalities and differences of the aforementioned examples with some unpublished observations conducted during a previous study. I conclude the lighting talk with a series of open research questions illustrated by other examples of people doing similar videos in various contexts.

3.17 Presentation-oriented vis techniques and how they differ from the usual analytical view

Robert Kosara (Tableau Software – Seattle, US)

Information Visualization is focused on the analysis of data, which shapes the techniques that get published and the criteria used to evaluate them. The goals for presentation are different, however. In particular, generality of techniques is not nearly as important, since the author of a presentation or story can try out different ideas more easily than during analysis. Using three examples, the connected scatterplot, ISOTYPE, and the StreamGraph, I argue that we need to understand techniques better that are useful for presentation and develop the tools to evaluate them.

3.18 Data Stories: Interactivity vs. linear storytelling – which way to go?

Ulrike Köppen (Bayerischer Rundfunk, DE)

Working in data journalism means to find answers to the questions: Do you want your users to lean back or to get involved in the story? Or both? And when exactly? In one of the last BR Data projects we designed an interactive player that combined explanatory and exploratory elements. The idea was to tell a linear story but to allow the user at certain story points to dig into the data (find the project here http://schnee-von-morgen.br.de/stage-1). Some of our takeaways: Don’t give the user choices right at the beginning Don‘t present the whole data set within a story … … but exactly those data sets that fit to the narration at that point

3.19 Classification of common data driving forces

Giuseppe Santucci (Sapienza University of Rome, IT)

Data driven stories use visualizations as building blocks for the story authoring. This talk presents a first classification of the most common data driving forces, i.e., how data characteristics are presented and used in the stories, like functions, proportions, and correlations. Moreover, it discusses how to formally express the mapping between the data and the abstract characteristics of the visualization. One possible outcome of this activity is the definition of a basic set of visualizations that can be used to quickly insert a 'visualization bricks' in the story during the authoring phase.

3.20 Responsible/ethical data visualization

Jonathan Schwabish (Urban Institute – Washington, US)

The changing value and availability of data has democratized people's ability to use data to conduct analysis, show results, and tell visual stories. But that same democratization opens the door for people to misunderstand and misuse data for their purposes. Using data responsibly and ethically is something all data creators and users must take seriously. More generally, presenting data responsibly may encompass a variety of different areas including risk and uncertainty, method and source transparency, data and visualization literacy, cultural, societal, and ethnic diversity of both users and creators.

3.21 Data-Driven Storytelling – An InfoVis/Computer Science Perspective

John T. Stasko (Georgia Institute of Technology – Atlanta, US)

In this talk, I present a perspective on the information visualization (InfoVis) research community by discussing its two main applications: analysis (exploration) and presentation (explanation). Until recently, research for analytic purposes has dominated the community. In particular, InfoVis thrives in situations of exploratory data analysis by employing clever techniques, multiple views, and interaction. Analysis benefits from visualizations that show all the variables, illustrate both overview and detail, and facilitate comparison. Such visualizations may thus require training and practice. Furthermore, visualizations designed for analysis emphasize completeness, accuracy/truth, flexibility and comprehensibility. This seminar focuses on the other application of InfoVis, its use for presentation purposes including storytelling. For this objective, people frequently design with goals of clarifying, focusing, highlighting, simplifying, and potentially persuading, and thus may not show all cases or attributes of the data. Additionally, somewhat in contrast to analysis, visualizations designed for presentation emphasize learnability, comprehension, usability, accuracy/truth, interest, and appeal.

3.22 Data Cuisine

Moritz Stefaner (Truth and Beauty – Lilienthal, DE)

Joint work of pozessagenten (Susanne Jaschko) and Moritz Stefaner
URL http://data-cuisine.net

Data Cuisine investigates the language of food to express information. It is a series of participatory workshops organized by Susanne Jaschko and Moritz Stefaner. In the workshops, we team up with a local chef and help participants to invent *data dishes* that represent local data in edible form. The dishes can vary in ingredients, taste, shape, color, temperature, texture, size, presentation, . . . to express information. The workshops end with a tasting and discussion of the developed dishes. We see "foodification" as form of data expression that might be imprecise, but on the other hand sensually intriguing, very association rich, and fosters debate around the data.

3.23 Visual Mementos

Alice Thudt (University of Calgary, CA)

In my talk I discuss the creation of visual mementos as a new application area for visualization. Visual mementos are visualizations of personally relevant data for the purpose of reminiscing, and sharing of life experiences. Today more people collect digital information about their life than ever before. The shift from physical to digital archives poses new challenges and opportunities for self-reflection and self-representation. Drawing on research on autobiographical memory and on the role of artifacts in reminiscing, we identified design challenges for visual mementos: mapping data to evoke familiarity, expressing subjectivity, and obscuring sensitive details for sharing. Visual mementos can make use of the known strengths of visualization in revealing patterns to show the familiar instead of the unexpected, and extend representational mappings beyond the objective to include the more subjective. To understand whether people's subjective views on their past can be reflected in a visual representation, we developed, deployed and studied a technology probe that exemplifies our concept of visual mementos. Our results show how reminiscing has been supported and reveal promising new directions for self-reflection and sharing through visual mementos of personal experiences.

3.24 Cognitive principles for visual explanations and narratives

Barbara Tversky (Stanford University, US)

Telling stories with data brings together many disparate strands of inquiry that have been studied in a variety of disciplines: ancient visual story-telling; perception and comprehension of the events from which stories are crafted, the production and creation of external representations of various kinds of information, the structures of stories, the understanding of stories, perception, comprehension, memory for various visual displays their use for understanding, inference, discovery, and action. Documenting what and how people create stories and represent data can provide cognitive principles for their design. Crafting visual stories of data depends on all that as well as the constraints of the media, notably print and digital, and the newsworthiness of the stories.

3.25 Exploring Data Sketching

Jagoda Walny (University of Calgary, CA)

Hand-drawn sketching is a common, accessible, and expressive method for generating visual data representations. We studied sketches of a single small dataset drawn by people with varying degrees of visualization experience, along with their reports of what they learned about the data. The sketches revealed a rich representation continuum from numeracy to abstraction; the data reports revealed a data report spectrum from individual data items to speculative data hypothesis. Relating the two results in strong potential directions for data-driven storytelling research.

3.26 Software Engineering in the newsroom

Stefan Wehrmeyer (Correctiv – Berlin, DE)

Software Engineering is present in all areas of media organisations: from the systems administration and IT-support (e.g. email system) to publishing content and tracking usage. Bigger media organisations that are well staffed with technologists can take matters further: they can better engage their audience, conduct investigations based on data and publish custom data visualisations and interactives. However, many older small to medium size media organisations have difficulties deploying these extra resources to take their publishing efforts further. Their constraints are financial, but they also suffer from accumulated technical debt or overreliance on outsourced services. Specifically their Content Management System acts as a gateway and restricts what they can publish. Additionally, journalism processes are focused on producing one-off articles and the editorial part of media organisation is often not used to the iterative nature of software development and the necessity for continued maintenance. Re-use of story-specific software between articles and extraction of their utility into libraries is difficult to fit into the publishing cycle. Many Free and Open Source software products have adopted a main-sponsor / multi-stakeholder approach. Smaller organisations can benefit from production-ready and maintained software without needing the same in-house development capacity as bigger organisations. A shared ownership approach for newsroom software and tools can support small and medium size media organisations that might otherwise struggle to survive.

3.27 Practical Process: An empirical study about visualization design workflows from around the world

Benjamin Wiederkehr (Interactive Things – Zürich, CH)

Visualization has proven to be a strong tool to explore and evaluate complex information. Let's discover how we can use data visualization as a way to tell engaging and compelling stories that are meaningful to your audience. Benjamin Wiederkehr, Managing Director of Interactive Things, shares insights how his team designs and develops visual and interactive stories based on quantitative and qualitative data. Based on a case study from their daily work, the long-form piece "Keine Zeit für Wut" for the Neue Zürcher Zeitung, Benjamin talks about the potentials and pitfalls of a structured process for telling data-driven stories.

3.28 Telling a data-driven story over a year, one day at a time: Cusum charts, cycling world records and community engagement

Jo Wood (City University – London, GB)

Since the early 1900s, cyclists have been telling data stories through their accounts and charts of miles covered by bicycle. Magazines invited readers to chart their weekly mileage and send in their charts for publication. This soon developed into a competition for the furthest ridden in a year. In 1939 Tommy Godwin broke that record with 75065 miles (120805 km). His progress in comparison to previous record holders can be shown as a cumulative distance chart. By rotating and rescaling such a chart, a stronger visual story can be told, in the form of a Cusum chart with the horizontal axis representing time, and the vertical representing distance above or behind a baseline. That baseline represents a goal in the form of a cumulative distance of another rider's progress for comparison. This has the advantage of showing local changes rapidly as they happen, while the smoothing effect of cumulative distance shows the bigger picture. This design was used in a community of cyclists to document the daily progress of five riders who challenged the 75 year old record in 2015/16. This form of chart is amenable to annotation, further increasing its story telling value. It was illustrated by showing the current record holder Kurt Searvogel's progress over the year.

4 Working groups

4.1 Evaluating Data-Driven Stories

Fereshteh Amini (University of Manitoba – Winnipeg, CA), Gordon Bolduan (Universität des Saarlandes, DE), Matthew Brehmer (University of British Columbia – Vancouver, CA), Christina Elmer (Spiegel Online GmbH – Hamburg, DE), and Benjamin Wiederkehr (Interactive Things – Zürich, CH)

The study of data-driven storytelling requires specific guidelines, metrics, and methodologies reflecting their different complex aspects. Evaluation is not only essential for researchers to learn about the quality of data-driven storytelling but also for editorial rooms in media and enterprises to justify the required resources the gathering, analyzing and presentation of data. The following framework takes the different perspectives of author, audience and publisher and their correspondent criteria into account. Furthermore it connects them with the methods and metrics to provide a roadmap for what and how to measure if these resulting data-driven stories met the goals. In addition, we explore and define the constraints which might limit the metrics and methods available making it difficult to reach the goals. We conclude with a graphic that gives an overview how these entities are linked to each other.

4.2 Devices and Gadgets for Data Storytelling: watches to augmented reality

Dominikus Baur (Andechs, DE), Tim Dwyer (Monash University – Caulfield, AU), Xaquin Gonzalez Veira (The Guardian Visuals – London, GB), and Bongshin Lee (Microsoft Research – Redmond, US)

This chapter will discuss the role of different hardware devices and media in visual data driven storytelling. The different form factors offer different affordances for data storytelling affecting their suitability to the different data storytelling settings, as identified by Lee et al. For example, wall displays are well suited to synchronous co-located presentation, while watches and virtual reality headsets work better for personal consumption of pre-authored data stories.

4.3 Ethics in data-driven visual storytelling

Gordon Bolduan (Universität des Saarlandes, DE), Nicholas Diakopoulos (University of Maryland – College Park, US), Marian Dörk (Fachhochschule Potsdam, DE), Kennedy Elliott (The Washington Post, US), and Xaquin Gonzalez Veira (The Guardian Visuals – London, GB)

Is the sample representative, have we thought of the bias of whoever collected or aggregated the data, can we extract that certain conclusion from the dataset, is it implying something the data doesn't cover, does the visual device, or the interaction, or the animation affect the interpretation that the audience can have of the story? Those are questions that anyone that has produced or edited a data-driven visual story has, or at least should have, been confronted with.

After introducing the space, and the reasons and implications of ethics in that space, the chapter will look at the risks, caveats, and considerations at every step of the process, from the collection/acquisition of the data, to the analysis, presentation, and publication, every point supported by an example of a successful or flawed ethical consideration.

4.4 Narrative Design Patterns for Data-Driven Storytelling

Jeremy Boy (NYU Polytechnic School of Engineering, US), Benjamin Bach (Microsoft Research – Inria Joint Centre, FR), Lyn Bartram (Simon Fraser University – Surrey, CA), Paolo Ciuccarelli (Polytechnic University of Milan, IT), Steven M. Drucker (Microsoft Research – Redmond, US), Yuri Engelhardt (University of Twente, NL), Ulrike Köppen (Bayerischer Rundfunk, DE), Moritz Stefaner (Truth and Beauty – Lilienthal, DE), Barbara Tversky (Stanford University, US), and Jo Wood (City University – London, GB)

In this chapter, we introduce a set of narrative design patterns for data-driven stories. These patterns aim to facilitate the creation of rich stories told with data visualization. There are many different ways to tell a same story. Emphasis can be put on certain events rather than others, characters can play more or less important roles, the implication (or voice) of the narrator can be more or less present. These variations usually have different effects on readers' emotions and/or understanding of the sequence of events told in the story. We identify a series of structural narrative patterns that can be used on their own or in combination to tell stories in an almost infinite number of ways. We present fifty of them here, and illustrate how they may help storytellers rethink the stories they want to tell. We assume that these storytellers already know what story they want to tell; who they want to tell it to; and why they want to tell it, i.e., what kind of effects and outcomes they want their stories to have. To frame our data-driven story design patterns, we also propose a threefold defining framework for data-drive stories, based on flow, data perspective, and framing (or context).

4.5 Audience

Kennedy Elliott (The Washington Post, US), Steven M. Drucker (Microsoft Research – Redmond, US), Samuel Huron (ENST – Paris, FR), Robert Kosara (Tableau Software – Seattle, US), and Jonathan Schwabish (Urban Institute – Washington, US)

Creators of data-driven visual stories want to be as effective as possible in communicating their message. By carefully considering the needs of their audience, content creators can help their readers better understand their content. We describe four separate characteristics of audience that creators should consider: expertise and familiarity with the topic, the medium, data, and data visualization; expectations about how and what the story will deliver; how the reader uses the interface such as reading, scrolling, or other interactivity; and demographic characteristics of the audience such as age, gender, education, and location. We then discuss how these audience goals match the goals of the creator, be it to inform, persuade, educate, or entertain. We also discuss certain risks creators should recognize, such as confusing or offending the reader, or using unfamiliar jargon or technological interfaces. To help support the chapter, we provide evidence from X case studies from a variety of fields including research, media, and government organizations.

4.6 Organizational Management

Jonathan Schwabish (Urban Institute – Washington, US), Christina Elmer (Spiegel Online GmbH – Hamburg, DE), and Benjamin Wiederkehr (Interactive Things – Zürich, CH)

Creating and producing data-driven stories requires a variety of personnel, skillsets, and tools. Be it in news organizations, technology or other private-sector firms, public or non-governmental organizations, and in the design and education fields, organizational structure will impact how a data-driven visual story is created, told, and published. In this chapter, we ask the important question of how organizations set up and run structures for visual storytelling? Are they hierarchical organizations with multiple levels of management, or are they flatter with multiple managers? Are content producers broken into separate groups or pulled together into broader teams? Are there common practices across these different sectors and what appear to be best practices and management structures?

To write this chapter, each author will speak to a handful of experts in the field of data-driven storytelling to see how they group and organize their production process. These interviews, as well as the existing literature on management organizations, will inform the basis of the chapter.

4.7 Exploration and Explanation in Data Driven Stories

Alice Thudt (University of Calgary, CA), Jason Dykes (City University – London, GB), Theresia Gschwandtner (TU Wien, AT), John T. Stasko (Georgia Institute of Technology – Atlanta, US), and Jagoda Walny (University of Calgary, CA)

In our group we discussed the difference between and integration of exploration and explanation in visual data-driven storytelling. Exploratory visualizations allow for a lot of freedom which can include changing the visual representation, the focus of what is being shown, and the sequence in which the data is viewed. They allow readers to find their own stories in the data. Explanatory stories include a focussed message which is usually more narrow and guides the reader usually in a linear way. We identified some advantages and some disadvantages of exploration and explanation as well as dimensions that help to describe and classify data-driven stories. We begin to describe the space by identifying freedom, guidance regarding representation, focus and sequence as well as interpretation as important dimensions of data-driven storytelling and characterize existing systems along these dimensions. As an outcome we aim to describe the space of exploration and explanation in data-driven stories in light of recent vis research and exploratory visualization and provide practical recommendations on ways in which we might integrate both aspects.

4.8 Techniques and Design Choices for Storytelling

Robert Kosara (Tableau Software – Seattle, US), Fereshteh Amini (University of Manitoba – Winnipeg, CA), Benjamin Bach (Microsoft Research – Inria Joint Centre, FR), Matthew Brehmer (University of British Columbia – Vancouver, CA), and Stefan Wehrmeyer (Correctiv – Berlin, DE)

We discuss techniques and design choices for visual storytelling grounded in a survey of over 60 examples collected from various online news sources and from KIIB award-winning visualization and infographic design work. These design choices represent a middle ground between low-level visualization and interaction techniques and high-level narrative devices or structures. We define several classes of design choices: embellishment, explanation, exploration, navigation, story presentation, emphasis, focus, and annotation. We provide examples from our survey for each class of design choices. Finally, we offer several case studies of examples from our survey that make use of multiple design choices.

4.9 From analysis to communication: supporting the lifecycle of a story

Melanie Tory (Tableau Software – Palo Alto, US), Fanny Chevalier (INRIA, FR), Marian Dörk (Fachhochschule Potsdam, DE), Jessica Hullman (University of Washington – Seattle, US), Bongshin Lee (Microsoft Research – Redmond, US), Giuseppe Santucci (Sapienza University of Rome, IT), Jarke J. van Wijk (TU Eindhoven, NL), and Stefan Wehrmeyer (Correctiv – Berlin, DE)

We will explore how tools can better support the authoring of rich and custom data stories, with natural / seamless workflows. We aim to understand the roles and limitations of analysis / authoring tools within current workflow practices and use these insights to suggest opportunities for future research and design.

We will first report a summary of interviews with practitioners; these interviews aim to understand current workflow practices for analysis and authoring, the tools used to support those practices, and pain points in those processes. We will then reflect on design implications that may improve tool support for the authoring process as well as research opportunities related to such tool support. A strong theme is the interplay between analytical and communicative phases during both creation and consumption of data driven stories.

Participants

- Fereshteh Amini
University of Manitoba – Winnipeg, CA
- Benjamin Bach
Microsoft Research – Inria Joint Centre, FR
- Lyn Bartram
Simon Fraser University – Surrey, CA
- Dominikus Baur
Andechs, DE
- Gordon Bolduan
Universität des Saarlandes, DE
- Jeremy Boy
NYU Polytechnic School of Engineering, US
- Matthew Brehmer
University of British Columbia – Vancouver, CA
- Sheelagh Carpendale
University of Calgary, CA
- Fanny Chevalier
INRIA, FR
- Paolo Ciuccarelli
Polytechnic University of Milan, IT
- Nicholas Diakopoulos
University of Maryland – College Park, US
- Marian Dörk
Fachhochschule Potsdam, DE
- Steven M. Drucker
Microsoft Research – Redmond, US
- Tim Dwyer
Monash University – Caulfield, AU
- Jason Dykes
City University – London, GB
- Kennedy Elliott
The Washington Post, US
- Christina Elmer
Spiegel Online GmbH – Hamburg, DE
- Yuri Engelhardt
University of Twente, NL
- Xaquin Gonzalez Veira
The Guardian Visuals – London, GB
- Theresia Gschwandtner
TU Wien, AT
- Nathalie Henry Riche
Microsoft Research – Redmond, US
- Jessica Hullman
University of Washington – Seattle, US
- Samuel Huron
ENST – Paris, FR
- Christophe Hurter
ENAC – Toulouse, FR
- Ulrike Köppen
Bayerischer Rundfunk, DE
- Robert Kosara
Tableau Software – Seattle, US
- Bongshin Lee
Microsoft Research – Redmond, US
- Giuseppe Santucci
Sapienza University of Rome, IT
- Jonathan Schwabish
Urban Institute – Washington, US
- John T. Stasko
Georgia Institute of Technology – Atlanta, US
- Moritz Stefaner
Truth & Beauty – Lilienthal, DE
- Alice Thudt
University of Calgary, CA
- Melanie Tory
Tableau Software – Palo Alto, US
- Barbara Tversky
Stanford University, US
- Jarke J. van Wijk
TU Eindhoven, NL
- Jagoda Walny
University of Calgary, CA
- Stefan Wehrmeyer
Correctiv – Berlin, DE
- Benjamin Wiederkehr
Interactive Things – Zürich, CH
- Jo Wood
City University – London, GB

Report from Dagstuhl Seminar 16062

Modeling and Analysis of Semiconductor Supply Chains

Edited by

Chen-Fu Chien[1], Hans Ehm[2], John Fowler[3], and Lars Mönch[4]

1 **Department of Industrial Engineering & Engineering Management, National Tsing Hua University, Hsinchu, Taiwan,** cfchien@mx.nthu.edu.tw
2 **Supply Chain Innovation, Infineon Technologies AG, München, Germany,** hans.ehm@infineon.com
3 **Department of Supply Chain Management, Arizona State University, Tempe, USA,** john.fowler@asu.edu
4 **Chair of Enterprise-wide Software Systems, University of Hagen, Germany,** lars.moench@fernuni-hagen.de

Abstract

In February 2016 the Dagstuhl Seminar 16062 explored the needs of the semiconductor industry for better planning and scheduling approaches at the supply chain level and the requirements for information systems to support the approaches. The seminar participants also spent time identifying the core elements of a conceptual reference model for planning and control of semiconductor manufacturing supply chains. This 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 appendices.

Seminar February 7–12, 2016 – http://www.dagstuhl.de/16062

1998 ACM Subject Classification B.2.2 Performance Analysis and Design Aids (Simulation), F.2.2 Nonnumerical Algorithms and Problems (Scheduling and Sequencing), H.4.2 Types of Systems (Decision Support, Logistics), I.2.8 Problem Solving, Control Methods, and Search (Heuristic methods, Plan execution, formation, and generation, Scheduling)

Keywords and phrases Modeling, Simulation, Supply Chain Management, Semiconductor Manufacturing

Digital Object Identifier 10.4230/DagRep.6.2.28

Edited in cooperation with Raphael Herding

1 Summary

Chen-Fu Chien
Hans Ehm
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, (typically) 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. The tested devices are then sent to regional warehouses or directly to customers. Wafer fabrication and probe are often called the front-end and assembly and test are called the back-end.

Supply chain management (SCM) problems have become more and more important in the last decade. This has been caused by the fact that front-end operations are often performed in highly industrialized nations, while back-end operations are typically carried out in countries where labor rates are cheaper. Moreover, there are centers of competencies (e.g. bumping) that may consist of only a few process steps that may be done in a different company owned facility or remotely by a subcontractor. These centers of competencies speed up innovations and reduce costs, but increase the complexity of SCM.

The semiconductor industry is capital intensive with the cost of an entire wafer fab up to nearly $10 billion US caused primarily by extremely expensive machines, some up to $100 million US each. 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. Capacity expansions are very expensive and time-consuming. This kind of decision is based on demand forecasts for the next years. Because of the rapidly changing environment, the demand is highly volatile. Consequently, the forecast is rarely accurate. The semiconductor industry is an extreme field for SCM 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 (cf. [1]).

The purpose of this seminar was to bring together researchers from different disciplines including information systems, computer science, industrial engineering, operations research, and supply chain management whose central interest is in modeling, analyzing, and designing complex and large-scale supply chains as in the semiconductor industry. Moreover, practitioners from the semiconductor industry who have frequently articulated their perception that academic research does not always address the real problems faced by the industry brought in their domain knowledge to make sure that progress towards applicability and feasibility would be made during this seminar. The seminar had 26 attendees from ten different countries (see participant list at the end of the report). We had participants from leading semiconductor companies Infineon Technologies and Intel Corp. as well as researchers who work closely with ST Microelectronics, Globalfoundries, and Taiwan Semiconductor Manufacturing Company (TSMC).

A primary purpose of the workshop was to extend the scope of the academic research community from single wafer fabs to the entire semiconductor supply chain. We show the principle architecture of the planning and control system of a semiconductor supply chain in Figure 1.

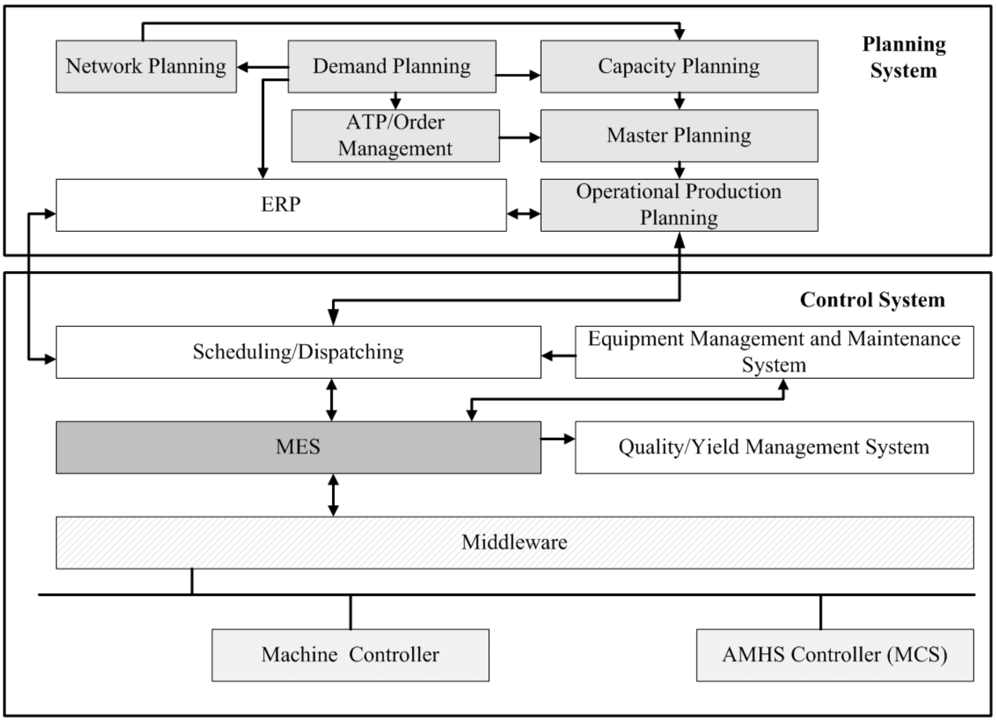

Figure 1 Planning and Control System of a Semiconductor Supply Chain (adapted from [2]).

Seminar Objectives

The first objective of the seminar consisted of developing a research agenda for semiconductor supply chain modeling and analysis topics. This includes innovative modeling approaches for supply chain network planning, demand planning, master planning, and detailed production planning and scheduling in semiconductor supply chains. But it also includes ideas on how to design the related future information systems.

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

- **Topic 1: Novel planning and scheduling approaches that can deal with the complexity and stochasticity of the semiconductor supply chain:**
  - Many planning approaches on the SC-level are based on (distributed) hierarchical and generally deterministic approaches to deal with the sheer complexity of the semiconductor supply chain. The role of anticipation of lower level behavior in upper level decision-making is still not well understood and has to be studied in more detail. Because a semiconductor supply chain contains many different, often autonomous decision-making entities including humans, negotiation approaches are typical in such distributed hierarchical systems for planning and control. It should be researched how such negotiation approaches can be automated and which decisions should be made by humans.
  - The overall cycle times in a typical semiconductor supply chain are on the order of 10 to 15 weeks. Therefore lead times have to be modeled in planning formulations. Using lead times as exogenous parameters in planning formulations leads to a well-known circularity because the cycle time depends in a nonlinear manner on the resource utilization which is a result of the release decisions made by the planning approach.

  Different types of clearing functions have to be researched in the semiconductor supply chain context.
  - Approaches to demand planning that take the product life cycle into account have to be studied. The interaction of demand planning and supply chain planning has to be investigated.
  - Different ways to anticipate stochasticity including robust optimization, approximate dynamic programming, and stochastic programming have to be researched in the semiconductor supply chain 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 data mining techniques have to be researched.
  - Because of the complexity of supply chains, long computing times still hinder the usage of analytic solution approaches especially for what-if analysis. 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 semiconductor supply chains has to be investigated.
- **Topic 2: Future information systems and supply chain management in the semiconductor industry:**
  - Understanding the limitations of today's packaged software for supply chain management in the semiconductor industry.
  - Proposing alternative software solutions including software agents and service-oriented computing for planning and scheduling applications in the supply chain context.
  - Integration concepts for state-of-the-art computing techniques to get models that are computationally tractable and address the different uncertainties encountered in this industry.
  - Approaches to embed real time simulation techniques in current and future information systems to support decision-making in semiconductor supply chains.
  - Understanding the interaction of human agents with information systems.

The implementation of ERP, APS, and MES systems in semiconductor supply chains provides both an opportunity and the need for development of supply-chain wide integrated production planning and scheduling solutions. Therefore, we think 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.

The second objective of the seminar consisted of identifying the core elements of a conceptual reference model for planning and control of a supply chain in the semiconductor industry that can be used for analysis and performance assessment purposes and to foster a common understanding in the research community both in academia and industry. This included specifying reference planning and control activities, the major information flows, and their interaction with a reference system of a physical supply chain. 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, mainly developed in the Measurement and Improvement of Manufacturing Capacity (MIMAC) project (led by one of the organizers of this Dagstuhl seminar) 20 years ago that are still used by many academic researchers working with the semiconductor industry.

Existing reference models on the planning and control level like the Supply Chain Operations (SCOR) reference model and the supply chain planning (SCP) matrix are too generic to be useful for detailed analysis and have to be refined considerably to cover the important domain-specific aspects of semiconductor supply chains.

The Process

In the opening session, the organizers welcomed the participants and acknowledged Infineon Technologies as a sponsor of the seminar. Next, the participants each introduced themselves. This was followed by an overview of the goals and objectives of the seminar and a detailed review of the seminar program including the ground rules for interactions.

The remainder of the day on Monday consisted of four industry overview talks (by Hans Ehm, Kenneth Fordyce, Chen-Fu Chien, and Irfan Ovacik) and a review of the literature related to modeling an analysis of semiconductor supply chains (by Lars Mönch and Reha Uzsoy). 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. See Table 1 for a list of topics and presenters and Section 3 for abstracts of the presentations.

Wednesday afternoon was the excursion that was enjoyed by the participants. Thursday was devoted to 3 breakout sessions with report outs on the topics in Table 2. Section 4 has the breakout report outs.

The first set of breakout sessions had four groups focus on the individual elements in Figure 1 and one group focus on a semiconductor supply chain reference model. The second set of breakouts had three groups consider the interaction between various elements in Figure 1, one group talked about the incorporation of humans in the supply chain, and one discussed how to go from the reference model to a specific semiconductor supply chain model instance.

The final Friday set of breakouts included three groups that discussed process models of multiple elements from Figure 1 and the flow of information needed between the elements to provide core elements of a reference model. Another group discussed the role of agents in a semiconductor company's supply chain. The final breakout group discussed the level of detail needed in a top down reference model. Friday consisted of a discussion on the required core elements of a reference model for semiconductor supply chains and a wrap-up session.

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 the different elements in Figure 1 are reasonably well understood by both the industrial and academic communities, but the interactions between the elements are less well understood. Having said this, a number of the software solutions for the elements are not geared toward the complexities of the semiconductor industry (e.g. ATP/APS systems are generally focused on profit maximization and ignore many of the system complexities). Second, it appears that there are still limitations in solution approaches in practice such as: capacity generally is expressed without regard to mix; fixed lead times are generally still assumed despite research done on clearing functions for planning; and ignoring all but production lots when developing

Table 1 Individual Presentations

Topic	Presenter
Network Planning Demand Planning Capacity Planning	Scott Mason Chen-Fu Chien Adar Kalir
Master Planning Order Release Planning ATP	Thomas Ponsignon Hubert Missbauer José Framinán
Global and Local Decisions Complexity in SCM Inventory Management	Stéphane Dauzère-Pérès Can Sun Jei-Zheng Wu
Semiconductor Supply Chain Contracts Supply Chain Planning Coordination via Planning Sustainability in SCM	Cathal Heavey Ton de Kok Jesus Jimenez
Simulation Modeling for Supply Chains (Distributed) Simulation for Semiconductor Manufacturing (Supply Chain) Decision Making Agent-based Simulation	Leon McGinnis Peter Lendermann Iris Lorscheid

Table 2 Breakout Sessions

Session	Topic	Participants (lead in bold)
1	Demand and Inventory Planning	**Uzsoy**, de Kok, Chien, Missbauer, Lee
	Capacity Planning	**Kalir**, Knopp, Lörscheid, Mönch, Dauzère-Pérès
	Master Planning	Fordyce, Herding, Mason, Ovacik, **Ponsignon**
	Available To Promise (ATP)	**Framinán**, Heavey, Ehm, Tirkel, Jimenez
	Reference Model	**Rose**, Sun, Weigert, Lendermann, McGinnis
2	Demand and Inventory Planning– Capacity Planning– Master Planning	de Kok, Tirkel, Uzsoy, **Dauzère-Pérès**
	Master Planning– Available to Promise	Framinan, **Fordyce**, Herding, Ponsignon, Kalir
	Master Planning– Factory	**Missbauer**, Weigert, Jimenez, Mönch, Knopp
	Incorporation of Human Behavior in the Supply Chain	Ehm, Heavey, Mason, Rose, **Lörscheid**
	From Reference Model to Systems	Chien, Sun, Lendermann, McGinnis, **Ovacik**
3	Process Models Demand and Inventory Planning– Capacity Planning– Master Planning	Kok, Tirkel, Uzsoy, **Dauzère-Pérès**
	Process Models Demand and Inventory Planning– ATP – Master Planning	Fordyce, Herding, Kalir, **Ponsignon**
	Process Models Master Planning– Factory	Missbauer, Weigert, **Jimenez**, Mönch, p
	Agents in the Level 3 Supply Chain	Lörscheid, **Mason**, Ovacik, Sun
	Top Down Reference Model	Ehm, Heavey, **Lendermann**, McGinnis, Rose

plans. Third, as indicated above both the industrial and academic participants generally agree that the integration of the decisions made by the different elements is often fairly ad hoc and could/should be improved. Finally, the participants generally agreed that there does not currently exist an adequate reference model for the semiconductor supply chain. In fact, there is not even a reasonable set of data sets that describe instances of the semiconductor supply chain such as the MIMAC datasets at the factory level. There is some indication that a reference model and incorporating human behavior of the various decision makers on the supply chain level will help to better understand supply chains producing and containing semiconductors.

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 because this is fundamentally more important than the actual solution techniques chosen. Some of the future research areas are included below:

- Using event-driven process chains (EPCs) to model/visualize planning processes.
- Developing better integration of various decisions made in the elements of Figure 1.
- Combining rolling horizon strategies with demand forecast evolution models.
- Incorporating sustainability aspects into supply chain models.
- Developing stochastic model versions of current deterministic models.
- Incorporating the behavior of human decision makers (this will be useful, but challenging).
- Exploring the use of different simulation paradigms (systems dynamics, agent-based, hybrid models, reduced simulation models) to model and analyze semiconductor supply chains.

Next Steps

As a way to further the discussion of and collaboration on the topics of the seminar, Prof. Lars Mönch, Prof. Chen-Fu Chien, Prof. Stéphane Dauzère-Pérès, Hans Ehm, and Prof. John Fowler are guest editing a special issue of the *International Journal of Production Research* (IJPR) entitled *Modeling and Analysis of Semiconductor Supply Chains*. The deadline for submission is September 1, 2016. This date was selected to allow time for ideas created by the participants of the seminar to be incorporated into papers for the special issue. The Call for Papers can be found at the following address:

http://explore.tandfonline.com/cfp/est/semiconductor-supply-chains-call

Acknowledgements. The seminar organizers would like to thank Infineon Technologies AG for their support of the 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., Dauzère-Pérès, S., Ehm, H., Fowler, J. W., Jiang, Z., Krishnaswamy, S., 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, 2011. http://dx.doi.org/10.1504/EJIE.2011.041616

2 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, 2013. ISBN 978-1-4899-9901-6.

2 Table of Contents

Breakout Reports

3 Overview of Talks

3.1 Industry Overview – Modeling and Analysis of Semiconductor Supply Chains

Hans Ehm (Infineon Technologies – München, DE)

The semiconductor innovation race continues, comparing with 6 years ago, many new technologies emerge, which also bring us a lot of challenges. The semiconductor supply chain is characterized by steep ramps, short product life cycles and long cycle times. Especially for companies who are far away from the end customers, the bullwhip effect is more significant and has a huge impact on the supply chain complexity. This speech discusses from Infineon, a company point of view and shares its best practices as well as its challenges on production planning and supply chain management. On one hand, Globalization and Flexibility are effective tools to manage global supply chain; on the other hand, the complexity of manufacturing process and planning is thus increased. The Modeling and simulation techniques become an important lever to solve complex problems and support decision making. A four-level approach is proposed: from the tool and work center bottom level to the end-to-end supply chain top level. A semiconductor supply chain simulation library called SCSC-SIMLIB was thus designed together with academia and implemented, and key objects in each level are developed. Following this structure many concrete examples and various applications within Infineon are given and demonstrated. Finally, the hot topics, e.g., human behaviors analysis, disruption management and change management are highlighted; and the future trends towards the Industry 4.0, IoT, and big data are referred to.

3.2 Supply Chain Management Planning for the Production of Semiconductor Based Packaged Goods Tasks, Purpose, Challenges, and a bit of History A Perspective from Agents of Change in the Trenches

Harpal Singh, John Milne, Ken Fordyce, and Robert Tenga (Arkieva – Wilmington, US)

Joint work of Peter Lyon, Gary Sullivan, Alfred Degobotse, Brian Denton, Bob Orzell

The purpose of any demand supply network is to meet prioritized demand on time without violating constraints and as much as possible meet business policies (inventory, preferred suppliers, request and commit date, etc.). Typically, the demand supply network for the production of semiconductor based packaged goods (SBPG) is divided into FAB and POST-FAB. The dynamic interaction between the two is limited in nature for logical and historical reasons. One reason is the nature of complexity which makes life interesting for planners is different between FAB and POST-FAB. FAB has long routes, reentrant flow, deployment and the ever present shadow of the operating curve – to name a few – generating wafer start / cycle time focus. POST-FAB is faced with constant exit demand uncertainty, allocation of shared components and capacity to competing demands, alternative operations; transport decisions, and the all-important "plan repair" – name a few – generating an exit demand /

efficiency frontier focus. Their differences become clear when examining the nature of the models (from spreadsheets to optimization) supporting decisions and analysis. The purpose of this presentation was to provide overview of the current best practices for central planning and identify computational challenges to reduce the current slack and make these networks more responsive.

3.3 Value Chain & Ecosystem Perspective for Modeling and Analysis of Semiconductor Manufacturing & SCM

Chen-Fu Chien (National Tsing Hua University, TW)

On the basis of my extensive collaborative studies with high-tech industries and the trends I have observed, this talk aims to address emerging research issues, from value chain & ecosystem perspective, driven by the needs in modeling and decision analysis in manufacturing that is enabled by the advances of automation and information technology. Semiconductor manufacturing is one of the most complicated and capital-intensive industries that is driven by Moore's Law for continuous improvements for technology advance and cost reduction. Therefore, high-tech companies are confronting with various decision problems involved in strategy, manufacturing, and technology that are characterized by uncertain (incomplete or massive) information and a need for tradeoff among different objectives and justification for the decisions to align with the overall strategic objectives. In the fully automation production facility such as semiconductor fab, manufacturing intelligence approaches have been developed to model the right decision problems involved in the operations and manufacturing strategies and to extract useful information to estimate the parameters and derive decision rules via data mining to enhance decision quality and production effectiveness. Owing to the advancement of information technology, researchers have developed various data mining methodologies to extract potentially useful patterns through semi-automatic exploration of a huge database in various domains. This talk will also use a number of empirical studies to illustrate the observation and the needs. Finally, we conclude this talk with discussion of the ongoing changes of manufacturing in high-tech industries and the emerging research directions. In addition, since the uncertainty involved in demand forecast is increasingly amplified with the forecast lead-time, high-tech companies often suffer the risks of oversupply and shortage of capacity that will affect the profitability and growth. High-tech industries including semiconductor and TFT-LCD industries are capital intensive, in which the capacity plan and corresponding capital investment decisions are critical due to demand fluctuation. Once the capacity is planned, the company may suffer the risks of either low capital-effectiveness due to low capacity utilization and capacity oversupply, or poor customer satisfaction caused by the capacity shortage. Most of the existing studies focused on solving the long-term capacity shortage issue through optimizing the capacity investment plan, or medium-term capacity plan to allocate demands among the wafer fabrication facilities (fabs) to balance the loading and product mix. Focusing on a real setting, this talk aims to share a proposed systematic decision method to analyze short-term solutions of cross-company capacity backup between the companies in the semiconductor industry ecosystem. In particular, a game theory and decision tree analysis model was developed to support this decision. A case study was conducted with real data of semiconductor manufacturing companies in Taiwan for validation. The results have demonstrated practical viability of this approach. The approach suggested has been implemented in this company. This talks concludes with a case study on the paradigm shifts of Global Unichip to address the issues involved in value chain & ecosystem perspective.

3.4 Master Production Scheduling Journey at Intel

Irfan Ovacik (Intel Corporation – Chandler, US)

In this talk, we discuss the journey that Intel took to build a world-class Master Production Scheduling (MPS) solution. MPS involves decisions as to what each factory in the supply network needs to manufacture, when, and at what quantities so as to meet the demand while minimizing supply chain costs. Most semiconductor companies have elected to use a best-in-class strategy and partnered with solution providers to build their supply chain planning solutions. In contrast, Intel chose to develop its solution in-house. We discuss how Intel decomposed the problem into smaller problems to better align with its strategy and organization boundaries and share the lessons learned during this journey.

3.5 Modeling and Analysis of Semiconductor Supply Chains: Preliminary Results of a Literature Survey

Lars Mönch (FernUniversität in Hagen, DE), Reha Uzsoy (North Carolina State University, US), and John Fowler (Arizona State University, US)

We conduct and report on a literature review of academic and industrial research in the domain of supply chain management in the semiconductor sector. Areas examined include demand planning, inventory management, network design, master planning, production planning, contracts and coordination, supply chain simulation, and involved Information systems. We also briefly discussed research from other domains and future research directions.

3.6 Network Planning

Scott J. Mason (Clemson University, US)

Semiconductor network planning is a challenging problem of great importance. However, not all plans are created equal – a variety of factors that can influence a plan include perspective, level of granularity, objective function(s), and whether or not uncertainty is included. We discuss an industrial case study using both Excel and mathematical optimization to analyze network plans for wafer starts, assembly starts, and test starts to minimize total costs.

We examine target inventory levels, lead times, and capacity constraints to illustrate key trade-offs faced by semiconductor manufacturers.

3.7 Modeling and Analysis of Semiconductor Manufacturing & SCM: Demand Forecast/Planning for Capacity Planning

Chen-Fu Chien (National Tsing Hua University, TW)

Semiconductor industry is capital intensive in which capacity utilization significantly affect the capital effectiveness and profitability of semiconductor manufacturing companies. Thus, demand forecasting provides critical input to support the decisions of capacity planning and the associated capital investments for capacity expansion that require long lead-time. However, the involved uncertainty in demand and the fluctuation of semiconductor supply chains make the present problem increasingly difficult due to diversifying product lines and shortening product life cycle in the consumer electronics era. Semiconductor companies must forecast future demand to provide the basis for supply chain strategic decisions including new fab construction, technology migration, capacity transformation and expansion, tool procurement, and outsourcing. Focused on realistic needs for manufacturing intelligence, this talk aims to share a proposed multi-generation diffusion model for semiconductor product demand forecast, namely the SMPRT model, incorporating seasonal factor (S), market growth rate (M), price (P), repeat purchases (R), technology substitution (T), in which the nonlinear least square method is employed for parameter estimation. An empirical study was conducted in a leading semiconductor foundry in Hsinchu Science Park and the results validated the practical viability of the proposed model. Furthermore, the forecasted demands can be used for capacity planning. Due to constant technology advance driven by Moore's Law in semiconductor industry, multiple production technologies generally co-exist in a wafer fabrication facility with utilization of a pool of common tools for multiple technologies and critical tools dedicated for a specific technology. In semiconductor industry, demand forecasts are rolling and updated when the latest market and demand information is available. This demand forecast mechanism makes forecast errors in different time periods correlated. Because part of the equipment is common for products of different technologies, production managers have limited flexibility to dynamically allocate the capacity among the technologies via capacity migration. The possibility of capacity migration and interrelationship among different technologies make capacity planning difficult under demand and product-mix uncertainties. We developed a dynamic optimization method that captures the unique characteristics of rolling demand forecast mechanism to solve capacity expansion and migration planning problems in semiconductor industry. We also proposed a mini-max regret strategy for capacity planning under risk. We estimate the validity and robustness of the proposed dynamic optimization method in an empirical study in a semiconductor manufacturing company in Taiwan. The results showed practical viability of this approach and the findings can provide useful guidelines for capacity planning process under rolling forecast mechanism.

3.8 Capacity Planning in Semiconductor Manufacturing: A Practical Perspective

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

Capacity planning problems have been researched extensively over the past two decades and still are, even with more intensity, in recent years. In this talk, an industry perspective is provided on the classification of the problems and the evolution in the various solution approaches to these problems. Insight to current and future challenges is also discussed with respect to both strategic and tactical capacity planning.

3.9 Master Planning in Semiconductor Supply Chains

Thomas Ponsignon (Infineon Technologies – München, DE)

Given the specifics of semiconductor manufacturing networks, the development of enterprise-wide planning approaches that are computationally tractable and address the uncertainties typically encountered in this industry remains particularly challenging. This talk focuses on Master Planning in semiconductor supply chains. Master Planning determines production quantities of end-products in a manufacturing network to meet external demands (e.g., customer orders and forecasts) and internal demands (e.g., requests for stock replenishment). A mid-term horizon (i.e., six months) expressed in weeks is considered. The outcome is capacitated production requests. Master Plan details the aggregated Sales & Operations Plan and it is the main input for Site Scheduling and Order Promising. In this presentation, Master Planning formulations and solving approaches as found in the scientific literature and the industry are discussed. Their performances with regard to solution quality, computational burden, and plan stability are outlined. Some insights from a semiconductor manufacturer into a real-world Master Planning approach are showed. Finally, trends and future challenges both for academia and practitioners are presented.

3.10 Optimization-based Order Release Planning

Hubert Missbauer (Universität Innsbruck, AT)

Order release is the interface between the centralized planning of the material flow through the entire logistic chain and detailed scheduling of the work orders within the production units. The presentation deals with multi-period models that optimize order release quantities per product and period based on a descriptive model of the material flow within the production unit that is represented as a network of work centers. We describe models with fixed target lead times as well as models with load-dependent lead times. Their shortcomings and the relevant research issues are outlined.

3.11 Available To Promise (ATP)

José M. Framinán (University of Sevilla, ES)

Available-To-Promise (ATP) systems deal with a number of managerial decisions related to Order Capture activities in a company, including order acceptance/rejection, due date setting, and resource scheduling. These different but interrelated decisions have been often studied in an isolated manner, and even the terminology and models employed differ largely. This communication intends to give an overview of the main contributions in the field and present some open issues for discussion.

3.12 Consistency between Global and Local Scheduling Decisions in Semiconductor Manufacturing

Stéphane Dauzère-Pérès (École des Mines de Saint-Etienne, FR)

Joint work of R. Sadeghi and S. Dauzère-Pérès

The operational level in semiconductor manufacturing can be divided into global (fab) and local (workshop) decision levels. The global level provides objectives or constraints for the local level. Based on previous research, our goal is to develop a framework to simultaneously optimize multiple performance measures and support the consistency between global and local scheduling decisions. A first approach is based on a global rule that aims at ensuring that lots satisfy their time constraints. Numerical experiments on industrial data show the interest of the approach. A first optimization model has also been proposed to take into account multiple constraints and objectives.

3.13 Complexity Management in Semiconductor Supply Chain

Can Sun (Infineon Technologies – München, DE)

Semiconductor supply chain is a complex system characterized by increasing number of interactive and interdependent components with dynamic behaviors working together as an entirety. Many innovative activities occur in the daily supply chain and manufacturing, and these changes inevitably bring in the complexities to the organization. But not all of them are valuable to the business goals. Decision makers want to keep value-added complexity and reduce non-value-added complexity. However, the quantitative analysis of the complexity generated by these components and their behaviors and its impact on the system still lacks practical methodology. Therefore we design a framework to measure the complexity of Semiconductor supply chain. The first step is to understand and represent the complexity using an conceptual model called PROS (process, role, object, state) idea, which provides an understandable and structural way to describe the complexity.., we can thus develop the formulas to measure system complexity based on the metrics of process complexity as well as the properties of complex system. A simplified small real example from semiconductor supply chain is used to demonstrate this approach. It is also noticed that the human plays an important role in the complexity due to its uncertainty behaviors. This is demonstrated and investigated through a classical supply chain phenomenon – the bullwhip effect, which can be demonstrated using a serious game called beer distribution game.

3.14 Inventory Management in Semiconductor Supply Chains

Jei-Zheng Wu (Soochow University, TW)

Joint work of C.-F. Chien, J.-Z. Wu, and H.-C. Yu

The rapid technology development and shortening product life cycle lead to high risk of product obsolescence. Manufacturers still need to hold a reasonable level of inventory to satisfy customers under demand uncertainty and long lead-time. This study introduces the inventory days into the multi-stage inventory model that incorporates with complex BOM, product substitution, wafer release schedules (wafer start), turnaround times (TAT, lead times), production plans, safety stock strategies, and end product-demand forecasts with four banks, i.e., VIA Bank, Wafer Bank, Die Bank, Finished-Good Bank for delayed differentiation and postponement strategy. Survival analysis is used to estimate inventory days and to group products. Inventory ages and accrued provision rates are also added into the model for the compliance with international financial reporting standards, No. 2.

References

1 Wu, J.-Z., Yu, H.-C., and Chien, C.-F. (2014/12), "Inventory survival analysis for semiconductor memory manufacturing," Proceedings of Winter Simulation Conference: Modeling and Analysis of Semiconductor Manufacturing (MASM), Savannah, pp. 2591–2599, GA, USA, December 7-10.
2 Wu, J.-Z., Chien, C.-F. and Tsou, Y.-C. (2014/08), "Multistage semiconductor memory inventory model based on survival analysis," IEEE International Conference on Automation Science and Engineering, pp. 613–618, Taipei, Taiwan, August 18-22.
3 Wu, J.-Z. (2013/05), "Inventory write-down prediction for semiconductor manufacturing considering inventory age, accounting principle, and product structure with real settings," Computers & Industrial Engineering, 65(1), pp. 128–136.

3.15 Semiconductor Supply Chain – Contracts

Cathal Heavey (University of Limerick, IE)

This presentation presented an introduction to contracts used in Semiconductor Supply Chain. It first presented the purpose of contracts, which are to coordinate a supply chain, the flow of product, information and funds. The presentation then stated that contracts need to be included into supply chain planning as they are a core element of planning in a similar way that inventory control is. The presentation also presented information on optimizing a RHF contract for a semiconductor SC with forecast error. It then presented an introduction to comparing option contracts with RHF contracts. The presentation finally stated that this is an important aspect of semiconductor SC.

3.16 Coordinating the Semiconductor Supply Chain by Planning

Ton de Kok (TU Eindhoven, NL)

Starting from the Eindhoven Framework for Production and Inventory Control (EFPIC) we discuss decision hierarchies and mathematical models used at different levels. The key feature of the hierarchy proposed is the (planned) lead time that enables decomposition between supply chain planning and production unit planning and scheduling. Though stochastic multi-item multi-echelon inventory models produce empirically valid results, they cannot cope with the planning complexities of the semiconductor supply chain. This leads to alternative mathematical programming formulations. We briefly discuss the interplay between APS and human planner and scheduler.

3.17 Sustainability in Supply Chain Management

Jesus Jimenez (Texas State University – San Marcos, US)

Joint work of Tongdan Jin
Main reference V. Santana-Viera, J.A. Jimenez, T. Jin, and J. Espiritu, "Implementing Factory Demand Response with On-site Renewable Energy: A Design-of-experiment Approach", in International Journal of Production Research – Special Issue on Energy-aware Manufacturing Operations, Vol. 53(23), pp. 7034–7048, 2014.
URL http://dx.doi.org/10.1080/00207543.2014.957877
Main reference S. Villarreal, J.A. Jimenez, T. Jin, and M. Cabrera-Rios, "Designing a Sustainable and Distributed Generation System for Semiconductor Wafer Fabs", in IEEE Transactions on Automation Science and Engineering, Vol. 10 (1), pp. 16–26, 2012.
URL http://dx.doi.org/10.1109/TASE.2012.2214438
Main reference L. Sanders, S. Lopez, G. Guzman, J.A. Jimenez, and T. Jin, "Simulation of a green wafer fab featuring solar photovoltaic technology and storage system", in Proc. of the 2012 Winter Simulation Conference (WSC), pp. 1–12, 2012.
URL http://dx.doi.org/10.1109/WSC.2012.6465269

The high-tech facilities used for the fabrication of semiconductor wafers consume a significant amount of electricity. The impact of energy consumption on climate change and the rising cost of energy have become a challenging issue facing the semiconductor manufacturing industry today. ITRS urges chip manufacturers to reduce carbon footprints by designing and deploying green and sustainable manufacturing facilities. The focus of this presentation is to present opportunities for the supply chain in semiconductor manufacturing. We studied the penetration of renewable technology in wafer fabs and identified the costs of these systems. We measure carbon emissions and probability of black outs. Future research is the development of models that measure carbon dioxide savings across the supply chain.

3.18 What Is a Reference Model and What Is It Good For?

Leon F. McGinnis (Georgia Institute of Technology, US)

Semiconductor supply chains are very complex, distributed, and dynamic systems, and it is simply not possible to make important decisions about designing, planning, or controlling them without computational decision support. Today, a grand challenge is to make that computational support as ubiquitous as using Google Maps on your mobile phone to find the short route from Frankfurt to Schloss Dagstuhl. In the status quo, however, these computational decision support tools require a great deal of customization for every different decision maker. So how do we change that?

An approach that holds great promise involves adapting two key concepts from computer science. The first is meta-modeling, or the creation of reference models of the domain of interest. These reference models can be articulated using languages like UML, its variant, SysML, or related Ecore, and define the semantics and to some extent the syntax for describing instances of problems in the semiconductor supply chain. The second is model-to-model transformation, which allows us to capture the "algorithm" for translating one model – say a formal model of a problem instance in the semiconductor supply chain domain – into another model – say a large scale optimization for planning the production and logistics for device manufacture over the next six months. Once these models are generated, there are many solvers that can be used effectively to compute solutions.

This presentation argues for the importance of meta-modeling and model-to-model transformation as keys for bringing academic research results into semiconductor supply chain practice faster, and more reliably.

3.19 (Distributed) Simulation for Semiconductor Manufacturing (Supply Chain) Decision-Making

Peter Lendermann (D-SIMLAB – Singapore, SG)

The presentation describes a number of challenges associated with the use of distributed simulation for decision-making in semiconductor supply chain management such as the heterogeneity of external drivers of such supply chains, the need of representing planning decisions and human decision-making as well as the constraints arising from time synchronization between the different federates of such a simulation system.

Additional challenges are arising from the fact that in a real fab environment product mix changes continuously and operations are never in steady state. To address this the author is advocating a simulation-based WIP Management approach, allowing a great reduction of variability in fab operations. One of the current research questions is whether this kind of variability reduction can also be reached on the Semiconductor Supply Chain level. This would obviously depend on which specific use cases are relevant on the Supply Chain level and which regular SCM decisions are to be enabled. As of now the within-echelon Semiconductor SCM challenge, i.e. the Borderless Fab, appears to be the most promising next application.

A number of key learnings have evolved from the author's experience with real-world semiconductor manufacturing systems and supply chains: In particular, there is no need to find an optimum because a real optimum can only be found with a perfect model. In reality it is sufficient to find a "considerably better" solution as fast (i.e. with as few iterations) as possible. This can be achieved through "smart" heuristics and enabled by a parallel computing infrastructure. It is essential though that constraints are portrayed as much as possible because otherwise "solutions" will be generated that are infeasible in practice and managers will lose trust in the enabling software tool. For this reason the Discrete Event Simulation approach can be quite powerful, however, automation is key not only for "analysis" but also for "modelling" including model verification and model maintenance as well as data calibration. Otherwise it would not be possible neither to keep up with the fast-changing real world operations nor to carry out "routine" analysis instantly and at zero (variable) cost.

3.20 Agent-based Simulation

Iris Lorscheid (TU Hamburg-Harburg, DE)

Agent-based simulation may provide new perspectives on supply chains from three angles: analyzing the supply chain as an emergent system resulting from individual interactions, incorporating the uncertainties caused by human behavior, and using artificial agents to find new designs or strategies. The first perspective models individual (inter-)actions in complex systems and analyzes the resulting system behavior. This provides an understanding on self-organizing processes and emergent phenomena that are not explicitly modeled or even understood by the modeler. By analyzing the human factor in supply chain planning processes, the second perspective, good or bad individual strategies may be observed, their effect analyzed, and the decision maker environment designed in a way that those are promoted or prevented. Artificial agents, finally, may help to optimize complex system by optimizing individual strategies by means of machine learning algorithms.

4 Breakout Reports

4.1 Demand and Inventory Planning

***Reha Uzsoy**, Ton de Kok, Chen-Fu Chien, Hubert Missbauer, and Peng-Chie Lee*

The purpose of demand and inventory planning is to create reliable forecasts of demand for the portfolio of products offered by a semiconductor manufacturer and to determine the amount of inventory needed to buffer against the inevitable demand uncertainty. Effective demand planning can guide companies to improve the accuracy of revenue forecasts, align inventory levels with peaks and troughs in demand, and enhance profitability. In today's environment, utilization is the "name of the game" since fab investment is the main cost component, but is mostly a sunk cost and has a long lead time to add capacity. Backend utilization is also important, but is generally not quite as important because the equipment is not as expensive and has shorter lead times to purchase. In addition, many companies subcontract backend operations. Overall, there are low marginal costs to meet demand and therefore profit margins are often relatively high for satisfying additional demand.

As seen in Figure 1, the output of demand planning provides inputs to both network planning (strategic needs) and capacity planning (tactical needs) in regards to likely future demand. It also provides inputs to the available to promise (ATP) system in terms demand expected in the near term. In some companies (such as TSMC), the demand planner uses data and models along with tacit knowledge to act as a middle man in consolidating regional sales plans, correcting for correlations between products, and in communicating with capacity planning.

There are a number of things that need to be done to improve the overall effectiveness of demand and inventory planning. First, it would be helpful to develop swim lane schemes of the relation between sales planning, demand planning and capacity planning. Second, even though utilization determines long-term planning, the short-term match of supply and demand requires new ATP functionality, in particular the need for better allocation schemes during tight demand is very important. Third, there is a need to better integrate production planning and inventory planning. As researchers investigate these issues, the focus should be on problem formulation, not on algorithm development and should keep in mind that determining the correct cost structure is key to arrive at the right solution. Finally, there is a need to distinguish between fabless companies, foundries and fab owners.

4.2 Capacity Planning

***Adar Kalir**, Sebastian Knopp, Iris Lörscheid, Lars Mönch, and Stéphane Dauzère-Pérès*

Capacity planning incorporates planning for future capacity to ensure that capacity is available for production planning (tactical) and adjusted for forecasted demand (strategic). Capacity planning is fed from the aggregate planning and feeds into the master planning. The following decisions are addressed by capacity planning:

- equipment changes (buy, convert, qualify)
- feasibility check for the production plan (capacity-wise, not schedule)
- wafer start quantities per period by product (aggregate, not fab level)
- performance expected (production cycle time)

Figure 2 shows where capacity planning is located in the production planning and scheduling hierarchy.

The following objective function is maximized:

- profit (revenue – cost)
- cost (wafer cost, unit cost, backlog and inventory cost)

Risk minimization is also considered. The different types of capacity planning problems are summarized in Figure 3.

The following input data is used for capacity planning:

- planning horizon and time bucket
- demand plan (after smoothing) per product over planning horizon
- current equipment, and capacity by fab

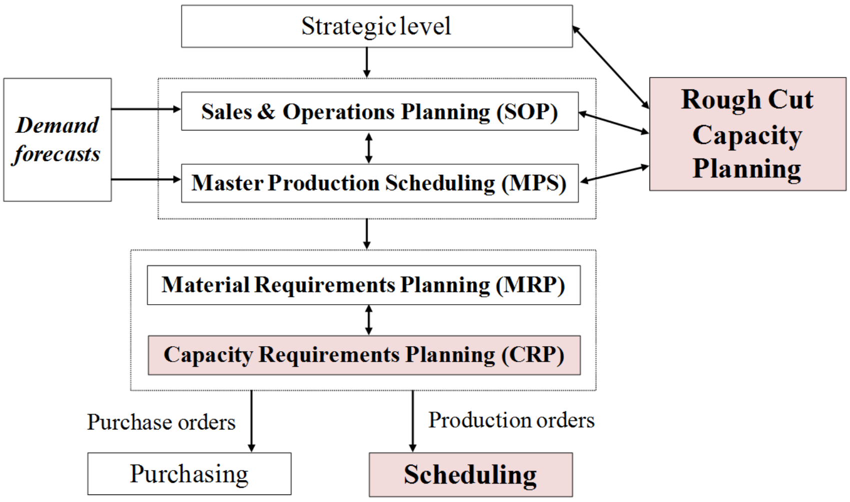

Figure 2 Capacity planning in the planning and scheduling hierarchy.

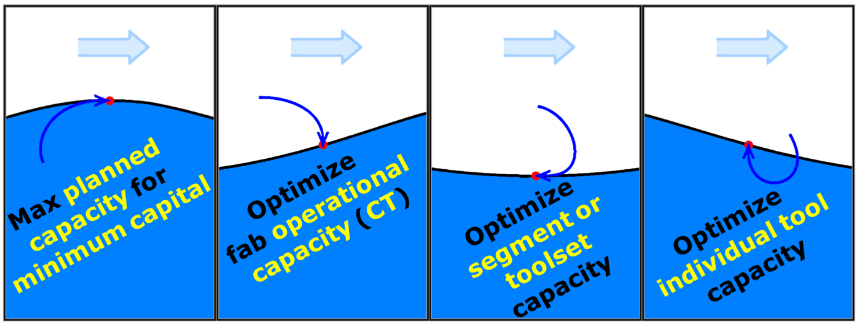

Figure 3 Mapping of Capacity Planning Problems.

- equipment data (RR, operations, ... – per product)
- capacity factors (dedications; setups; batching; ...)
- financial data (equipment value, wafer cost per fab)
- budget for expansion

The interface with network/demand planning is demand forecasting by product over time. The impact of multi-fab capacity is not considered. Typical modeling types are Excel for static considerations, MILP for deterministic settings, and simulation and stochastic programming for a stochastic setting.

Next, limitations of the current state and future needs are discussed. The hierarchy enforces 'independent' problems. However, capacity planning must be considered in the network/demand planning phase. Another imitation is the limited model accuracy because of time bucketing, the estimation of operational decisions such as setup, batching, QT, storage and transportation (especially for older manual fabs), reticles, etc., the deterministic nature of the forecast (point forecasts vs. the real dynamic nature of actual demand) and sensitivity to the impact of product mix. Current approaches are limited to production lots, often ignoring all other type of non productive wafer (NPW), engineering activities, new product introduction (NPI). The quality of the input data (typically aggregated) is crucial, this is especially true for scenarios for stochastic planning. APC/AEC and dedication schemes are not captured at all yet. Capacity planning is not adapted to the evolving SC in semiconductor industry. The dynamics of changing production across fab locations, cross-processing between fab locations, etc is not considered. A consideration of the impact of multi-fab capacity is important. In future modeling, agent-based simulation, game theory, and nonlinear programming are important ingredients for addressing new complexities. Moreover, rolling horizon approach are expected to be desirable to model decision-making changing over time.

The following issues have to be addressed in future research. From an industry point of view, considering NPW, engineering activities, and NPI in capacity planning is important. Moreover, getting refined model accuracy is desirable. The impact of multi-fab capacity has to be considered. From an academic point of view, getting more accurate models is one goal. Rolling horizon approach, capacity planning for the evolving SC, and considering the impact of multi-fab capacity are important future directions for academic research.

4.3 Master Planning

Kenneth Fordyce, Raphael Herding, Scott Mason, Irfan Ovacik, and ***Thomas Ponsignon***

Definition and Scope

Master Planning determines production quantities of end-products in a manufacturing network to meet external demands (e.g., orders and forecasts) and internal demands (e.g., requests for stock replenishment). A mid-term horizon (i.e., six months) expressed in weeks is considered. The outcome is capacitated production requests. The master plan details the sales & operations plan and it is the main input for site scheduling and order promising.

Current state-of-the-art

Most semiconductor manufacturing companies blend commercial solutions (e.g., JDA) and home-grown solutions (e.g., Intel). Regardless of the approach, most firms have operations research experts for development and analysis. The solution approaches typically involve solvers, which may be based on optimization techniques and/or rules-based heuristic approaches (e.g., forward-and-backward demand-supply assignment). A combination of optimization techniques and heuristic methods is frequently used in the industry. Software developers are key for the implementation and the deployment within a company (e.g., UI, data integration).

Inputs to Master Planning

Master planning requires the following inputs: the structure of the supply chain, demand signal, information on capacity, current inventory and work-in-progress levels, safety stock targets and policies, and the existing or previously generated master plan. The structure of the supply chain includes details on which products can be produced where, bill-of-material information, geographical information, details on available equipment as well as planning parameters such as lead times and planned yields. The demand signal is expressed as point estimates and capacity is often stated in gross terms (e.g., total starts per technology per week, hours of loading on a tool group).

Outputs from Master Planning

Master planning provides expectations or targets of what is required from each factory over time. It also states what each factory is expected to receive over time and from where. These outputs can be summarized as the current projected supply, which is provided to Available-to-Promise (ATP) usage and production planning processes. In some firms, the master plan may be a starting point for manual adjustments.

Limitations

Capacity statements are often given without consideration of mix. Furthermore, no common standard is available for describing capacity. The assumption of fixed lead times are here to stay as they are often politically influenced. This leads to input factors being conservatively stated. Deterministic point estimates do not capture the uncertainty of the base system. Hence, safety stock buffers for yield and demand variations may be overstated. As a result, potential opportunities may be lost. The interpretability remains a major limitation especially when optimization techniques are used to solve the master planning problem. Besides, little is known about the recommended plan granularity and model accuracy (e.g., time bucket specification).

Future steps

Directions for further research include: investigating which solution procedure works best in which situation, examining the usage of blended optimization and heuristic solution methods, Incorporating risk assessment into master planning models, and tightening model integration with data sources and other models.

4.4 ATP

José Framinán, *Cathal Heavey, Hans Ehm, Israel Tirkel, and Jesus Jimenez*

ATP it is all about confirming customer orders. The product the customer/channel is ordering determines the granularity. A "contract" with customer, the Operations, one initial negotiated price, and the heuristic of the implemented APS are prerequisites.

Continuously updated capacity, inventory, committed forecast/order, requests for confirmation are important input data. Here, capacity means supply in a two step process where supply is generated from capacity.

The following activities are required:

- check Inventory
- check Supply
- use Alternatives
- continuous assessment of confirmation process.

Response to forecast/order and rigger for capacity increase are outputs. Next, the current situation in companies is discussed. ATP usage systems (most of the time called APS and developed outside the semiconductor industry) are in place. Online data for the ATP usage system is already available at some semiconductor companies. ATP usage systems used in

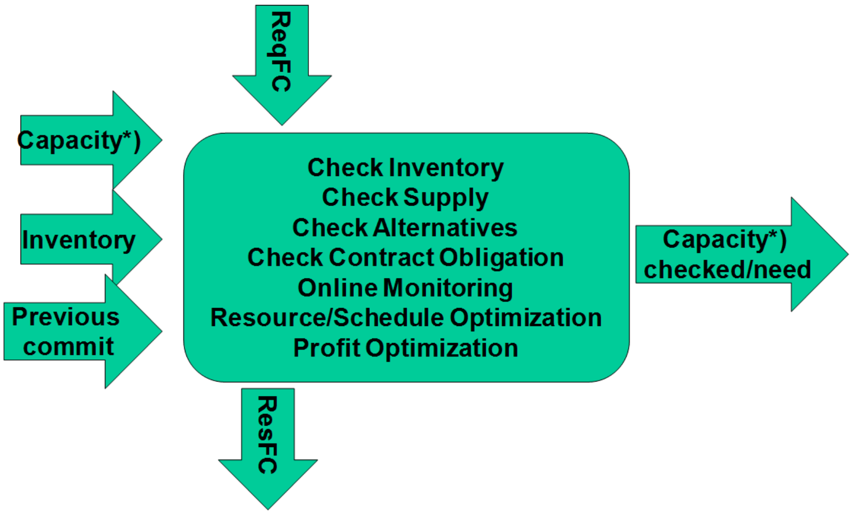

Figure 4 ATP Usage System.

the semiconductor industry are far from optimum. APS systems are optimized towards profit optimization and lack capacity/supply complexity need of the semiconductor industry. The full usage of resources is key for semiconductor companies. The main functionality of an ATP usage system is summarized in Figure 4.

The following future needs are identified:

- Make all data needed online available for an ATP usage system (at some companies available).
- Optimize the resources, and optimize the schedule (the heuristics in the ATP usage system).
- Make use out of the profit optimization in available ATP usage systems, i.e. short term benefit, mid term benefit, long term benefit.
- Measure the benefit.
- Measure the stability of the system.

4.5 Reference Model

Oliver Rose, *Can Sun, Gerald Weigert, Peter Lendermann, and Leon McGinnis*

The question what is a reference model is discussed first. The MIMAC data sets are instances that conform to some (unstated) reference model. A domain specific language (DSL) conforms to a reference model. A DSL is specified by symbols and rules and has a well-defined semantics. SCOR is "part of" a reference model, however, there are possible issues with precision of semantics definitions.

Next, the state of the art for supply chains is discussed. SCOR is more like an agreement how to present KPIs, it has a very simple process model. SAP (ARIS), Oracle are used for company-specific SC software design, but is not public. There exist various custom models.

Very general approaches are offered by OMG that are mainly from/for computer science. There are tools such as MOF, UML, SysML, BPML.

The following limitations exist. The motivation for developing a reference model is with respect to ROI not clear. Appropriate tools and tool chains do not exist or have some limitations. Some inertia with respect to reference modeling can be observed in organizations. The awareness of reference modeling issues by possible stakeholders is limited.

A SCSC DSL and tools (mock-ups) are identified as possible output of future research activities. This includes appropriate demos. The focus is on the internal supply chain (third level). Applying for Horizon 2020 projects (maybe only as a hidden agenda of a project) might be a future step. A core working group/task force (driven by industry needs) is highly desirable. A conference track or a journal special issue can be organized.

4.6 Demand & Inventory Planning – Capacity Planning – Master Planning

Ton de Kok, Israel Tirkel, Reha Uzsoy, and ***Stéphane Dauzère-Pérès***

Figure 5 shows the connections between the three planning modules, where capacity planning is actually divided into medium-term and long-term capacity planning.

Table 3 specifies various elements that characterize the planning modules.

Some comments are given below that are important in understanding how these planning decisions are performed:

- The role definition (in particular incentives), behavior and expertise of planners strongly impact the decision process,
- Delays (to get information and take decisions) are often non negligible (hours or days) and prevent the decision process to be conducted in real time,
- The informal communication among planners, sales, production and customers is critical to take informative and acceptable decisions.

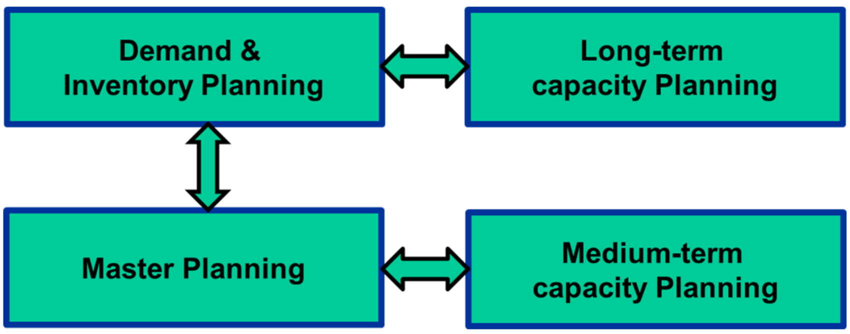

Figure 5 Relationships between planning modules.

Table 3

	D&I Planning	Capacity Planning		Master Planning
Input	Forecasts and orders	**Long term**: S&OP, aggregate capacity parameters	**Medium term**: MPS, detailed capacity parameters	Aggregate plan (S&OP) and orders
Output	Aggregate plan (S&OP), safety stocks	Projected capacity plan	Projected capacity plan	Master Production Schedule
Feedback loops	With Sales, with Capacity planning and with Master Planning	With D&I Planning and medium-term capacity planning	With Master Planning and Production (PP&S)	With everybody...
Granularity	One year or more with monthly / weekly buckets	One year or more with monthly / weekly buckets	6 months with weekly / daily buckets	6 months with weekly / daily buckets
Frequency	Monthly	Monthly	Weekly	Weekly
Tools	Multi-echelon inventory models, Statistical and qualitative forecasting	Stochastic models, LP models, MILP models	Stochastic models, LP models, MILP models, heuristics	LP models, MILP models, heuristics
Simulation	Appropriate at all levels in different ways			

4.7 MP & ATP

*José Framinán, **Ken Fordyce**, Raphael Herding, Thomas Ponsignon, and Adar Kalir*

The Central Planning Process (CPP) to manage the end to end demand supply network (DSN) for the production of semiconductor based packaged goods (SBPG) is focused on understanding and capturing exit demand, intelligently matching assets (WIP, inventory, and capacity across the network) to create a projected supply line linked to demand and synchronizing (but not scheduling) activities across the DSN.

Figure 6 has a high level summary of the core components in CPP and their relationship. The "order handling activity" (where ATP occurs) captures the requests of products (orders) from customers. An incoming order is either accepted and given a commit date or identified as pending further review. Orders (accepted and pending) are sent to the "demand and inventory planning activity" where a comprehensive statement of exit demand is established. This demand statement is sent to the "master planning activity" which matches assets with demand to create a projected supply line linked to demand. This is the basis for the commit decision – that is what product, in what quantity, can be committed to the customer or made available for the 'automatic' ATP process – that is when a commitment is made without going through the master planning process.

Although there are differences between firms, this is the basic process. Potential areas of

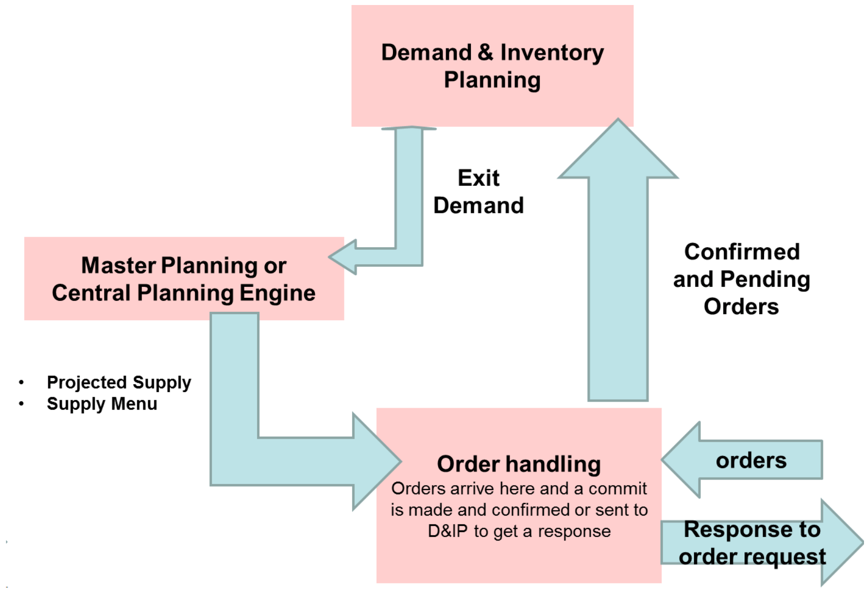

Figure 6 Summary of Master Planning and ATP flow.

improvement that will positively impact a firms' responsiveness are:

1. Improved flexibility in the projected supply used by the ATP, especially with regards to level of personalization. For example, assume I make T shirts in three colors: red, blue, and white – where red and blue shirts are made from white shirts going through an end step of dying and heating. I might project 100 reds, 100 blues, and 200 white as supply, but until the dye is done, this can be modified. Going a step back to material, until the shirt is cut, I have options to convert large to small, small to medium, etc. Often this flexibility is only captured if the in the master planning or central planning engine activity
2. On the order receiving side – being able to capture the "confidence"that the customer order is firm and using that within planning process would be most beneficial – that is how to incorporate uncertainty without creating confusion.
3. Faster and more intelligence central planning engines would be very beneficial – creating tighter coupling.

Figure 7 provides a more detailed view of the key components in master planning and ATP to create a projected supply linked to demand and synchronize the activities of the demand supply network (DSN), and planners workbench. From an operations research perspective, the most significant technical achievements occurred in the CPE (Denton et al., 2006 and Degbotse et al., 2013). Fordyce et al. (2011) has the best overall description of Order Planning System (OPS). The business contribution of OPS is covered in Lyon et al. (2000), Denton et al., (2006); and Fordyce (1998 and 2001). The work done by IBM in collaboration with Arkieva on demand management captured a critical paradigm shift that successful demand management was about a single integrated and flexible view of the key data sources (orders, forecast history, shipping, etc.), capturing sales estimates, collaboration, and developing insight – where statistical forecasting methods was just one component (Fordyce and Sullivan, 2016); this approach has carried forward to become best practice. Today the features and functions in OPS are standard best practice.

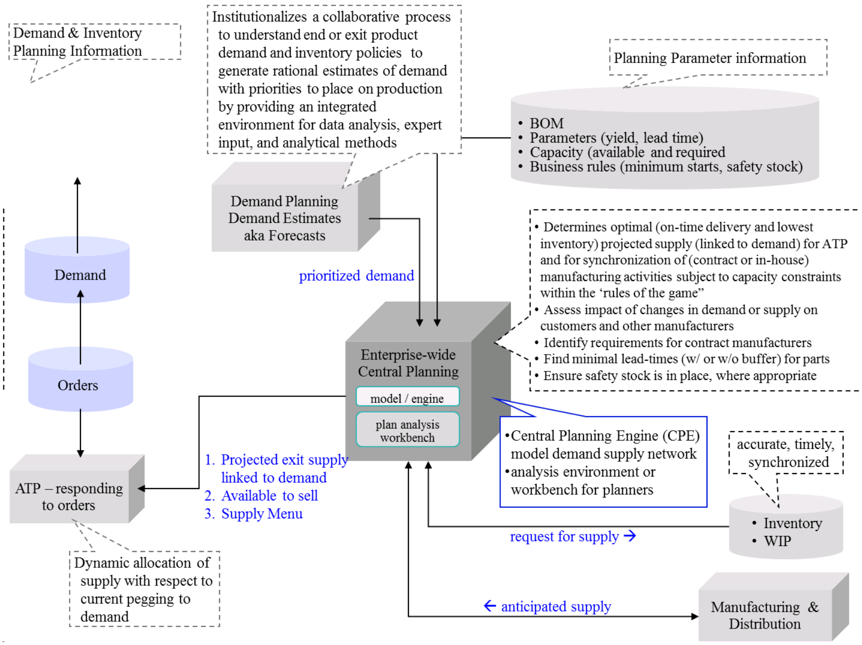

Figure 7 Detailed Components and Flow for Master Planning and ATP.

4.8 MP & Factory

***Hubert Missbauer**, Gerald Weigert, Jesus Jimenez, Lars Mönch, and Sebastian Knopp*

Interface to Master Planning (MP)

"Master Planning provides expectations or targets of what is required from each factory over time." (Master Planning – Breakout I). That is, quantities to be finished per product, period (usually weeks) and production unit (e.g., Fab, A/T, DC). Also the allocation to demand classes (e.g., customer order driven vs. forecast driven) which might affect the priorities.

The subproblems in order to finally arrive at a satisfactory schedule are the following:

- Split up the MP quantities into production lots, defined in terms of product, lot size and required due date in days. Both backward scheduling and forward scheduling should be possible.
- Order release (planning), where an order is a production lot. The extent to which releases are planned for several periods depends on the applied method (short-term release mechanisms vs. multi-period release planning). At this point the relationship between required output, target WIP and flow time norms must be considered. The underlying model of the material flow through, e.g., the fab is crucial. Target flow times can be fixed or load-dependent (see presentation Hubert Missbauer). Flow time anticipation using flow factors assumes strong correlation between processing time and flow time of an operation which is not always given. Order release usually is performed periodically. This can be complemented by event-driven releases (e.g., if otherwise machine idleness would occur), depending on the release mechanism.
- Due to the complexity of the material flow within a fab a global scheduling level ("Lot Planning") is required – setting intermediate due dates for production phases (work

centers or groups of work centers). This requires an overview of the whole fab which the local dispatchers do not have. (See presentation Stéphane Dauzère-Pérès.) End-of-horizon effects can occur, this requires look-ahead-feature.
- Based on the intermediate due dates for the released lots, detailed scheduling – dispatching is performed locally per production phase. This is often rule-based and requires extensive knowledge about the production process at this stage.
- Feedback (MES): The state of lots and work centers must be known at each time. This feedback information is transferred to the respective decision levels. In any case this must be the release level since load-based release requires accurate state information. Which decision level should react to unplanned events (e.g., machine downtimes) is a difficult question; as a general rule it should be the lowest possible level. Minor disturbances will just affect the dispatching level whereas major machine downtimes might even affect capacity and master planning.

Additional remarks

There are groups of wafers that share the first n operations; sometimes these base wafers are stored in order to benefit from variability pooling. This is an adaptation to individualization of customer needs. We discussed if it makes sense to formalize this and to declare the base wafers as SKUs (which implies a second release decision at this point).

Lot-to-order matching is an important task. How to assign this task to planning levels requires further discussion.

4.9 Incorporation of Human Behavior in the Supply Chain

*Hans Ehm, Cathal Heavey, Scott J. Mason, Oliver Rose, and **Iris Lorscheid***

The aim of this seminar is to provide a reference model that describes the structure, processes, and basic skeleton of a supply chain in semiconductor industries. However, this reference model not includes a relevant complication, which follows from the uncertainty of human behavior within supply chains. Thus, complementing to the reference model, we aim for incorporating human behavior by exploring ways to identify relevant individual behavior and understand their effect on supply chain performance.

First, the identification of individual strategies clarifies the existing individual deviations. Also, we aim to reveal (informal) interactions that proofed to support the quality of decisions and were therefore implemented by individuals. A simulation model may analyze circumstances that extend rational behavior, and aspects that may hinder rational behavior. Simulation experiments may quantify the risks for supply chain performance caused by less rational behavior.

Overall, we aim for improving the circumstances to support good decisions of planners. Knowing about favorable individual behavior, we may relate this to the right level of education that is required for the decision makers. Results can define criteria for recruiting with regard to technical and soft skills, and identify training requirements. By learning about the effect of varying strategies, the awareness and sensitivity to acknowledge (un-)favorable aspects of individual behavior can improve. In the end, we aim to avoid negative human subjectivity.

The first important step is to understand what humans actually do in our supply chains to understand the relevance: Where is the "Ron" in the model? As we know, Ron may have individual interests, for which he under- and over-estimates, and thus potentially behaves rational in his own interest but not rational for the company. In particular within demand, capacity, or master planning we observe persons who may "pre-process" data by adjusting regulation screws that improves individual planning result, such as by adapting demand, capacity, yield, or – as common example – lead time. We see events where undesirable behavior happens, and cases with undesired consequences. Also extreme cases such as hyper-optimistic or hyper-conservative planners are observable. Next to the planning scenario, other perspectives on the supply chain incorporate human behavior, such as negotiations between sales persons and customers and their effect on (un-)successful pricing, and order behavior by customers resulting in varying demand patterns.

We see some challenges for incorporating human behavior. First, the design of incentives can be relevant, so that individuals behave in the interest of the company rather than choosing fast and easy moves when fulfilling their tasks. Nevertheless, the design and successful implementation of incentives is challenging. Next to unfavorable behavior, we know about individual strategies that really stabilizes the supply chain. In compliance, for example, individual adaptations are necessary to fulfill the compliance challenge on the one hand, but be right for the internal processes such as product planning on the other hand. Thus, we know that individual behavior may work in certain circumstances but may lead to damage in others. The question is: How many Rons (or "Hans") do we need to keep the system running?

A further challenge with regard to collecting data for the identification of human behavior can be the identification of the "real" human factor that leads to the respective individual strategy of interest. Dominant factors such as experience, age, or seniority may prevail other aspects.

As concrete next step we propose to create a list of individuals that have an impact on the supply chain, with a description of why and how they have an impact. The definition of cases with positive and negative effects of human interactions on supply chain performance may clarify the relevance of incorporating human behavior. We expect that positive cases of individual strategies will include situations comprising complications such as innovations or complex tasks. Also, a literature review about similar studies about the effect of human factors on performance may support the research process.

4.10 From Reference Model to Systems (the SC System in 20201)

Chen-Fu Chien, Can Sun, Peter Lendermann, Leon F. McGinnis, and ***Irfan Ovacik***

The Reference Model for the Semiconductor Supply Chain contains a set of objects and the relationships between those objects, along with behaviors and controls that describe how the Semiconductor Supply Chain changes/evolves over time. These objects, behaviors and controls are described using a Domain Specific Language.

The Reference Model can then be used to instantiate/describe the physical supply chain of any semiconductor company. We refer to this as the Model Instance.

In academic settings and in practice, we usually talk about Math Programming, Queuing, or Simulation Models. In practical settings, we also find these models embedded in planning applications, wrapped with system and user interfaces. We will refer to these as Analysis Models. Each Analysis Model consists of inputs, algorithms, and outputs and targets a specific business decision.

Today, each Analysis Model potentially assumes a different Reference Model. For example, a JDA application may assume a different Reference Model than an SAP module even though each serves the same business function. Therefore, there is no guarantee that an Analysis Model that works in one setting will work in a different setting without considerable effort. Similarly, there may be considerable different effort to implement a JDA application than an SAP module in the same company.

In the long term, once a Reference Model for the Semiconductor Supply Chain is in place, then each Analysis Model in the environment is expected to conform to the Reference Model. Each Analysis Model can use a subset of information available in the Reference Model or an abstraction of it. For example, a Production Planning and Scheduling model would only use the relevant information for the Back End Supply Chain or a Master Planning model would treat each factory as an abstract black box, ignoring the details of each factory. In both cases, the Analysis Model would be consistent with the Reference Model. To the extent that the Analysis Model conforms to the Reference Model, then the model would be reusable whether it is interfacing with a Model Instance associated with company A or the Model Instance associated with Company B.

4.11 Process Models Demand and Inventory Planning, Capacity Planning, Master Planning

Ton de Kok, Israel Tirkel, Reha Uzsoy, and Stéphane Dauzère-Pérès

The relations between demand planning, inventory planning, capacity planning and master planning

We considered the recurrent communication between Demand and Inventory Planning (DIP), Capacity Planning (CP) and Master Planning (MP) at monthly and weekly frequency. Starting with monthly regional sales forecasts, which are consolidated into product forecasts by a demand planner, DIP communicates monthly with CP to agree on a long-term capacity and sales plan. This process is iterative as capacity may not be able to accommodate the initial sales plan. This results into updated regional sales plans which are input for MP. We realized that our process is closely related to the TSMC process, which focuses on wafer production. Companies with an integrated front-end and back-end start from customer sales plans at SKU level, which is consolidated with respect to resource requirements and possibly to SKU requirements in case multiple customers need the same IC. But similarly the long-term consolidated sales plans must be translated into long-term capacity requirements by capacity planning. A similar iterative process leads to updated customer (or regional) sales plans, which are input for MP.

MP processes the sales plans and derives time-phased key component requirements and key resource requirements. The latter are discussed with CP to check if mid-term the required resources are available. Again, an iterative process yields an updated MP, which is

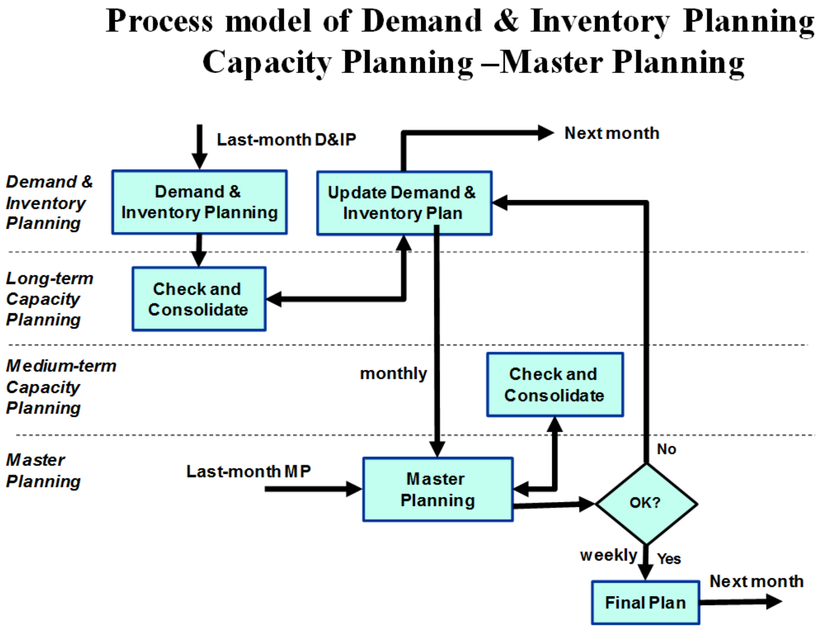

Figure 8 Process Modell Demand & Inventory Planning, Capacity Planning, Master Planning.

communicated with sales. This may lead to mid-term adjustments to the sales plans. This process repeats itself weekly.

We closed the loop by using the sales plans and masterplans of the current month as benchmarks for the sales planning and capacity planning for the next month.

4.12 Process Models Demand and Inventory Planning, ATP, Master Planning

Kenneth Fordyce, Raphael Herding, Adar Kalir, and ***Thomas Ponsignon***

Event-driven process chains Demand Planning – Master Planning – ATP Usage

New orders enter via the Order Handling System. A new order entry triggers the provision of confirmed and pending orders to Demand Planning. Demand Planning involves collaborative process/human activities. It is typically run according to time-scheduled events. The output from Demand Planning is a Demand Outlook, which is provided to Master Planning. Master Planning accepts further inputs from Capacity Planning and Production Management that are not represented on this chart. Master Planning involves a solver. The output is a projected supply or a supply menu (i.e., available supply with a set of conditions such as pricing and given lead times) depending on the business model of the company. The supply picture is provided to the ATP Usage, which involves an Order Promising System. The ATP Usage activity is usually run according to an order fulfilment event. It results an order response that is sent to the customer. Further information may be added to the process flow

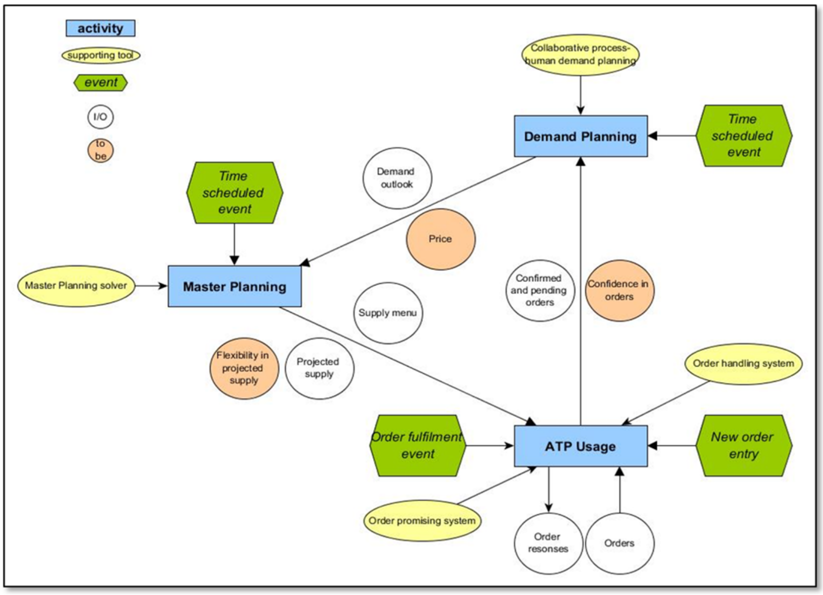

Figure 9 Process Model Demand Planning – Master Planning – ATP Usage.

to facilitate the decision-making: confidence in orders may help to capture the uncertainty of demand; price may support the prioritization of demand items in Master Planning; and Flexibility of projected supply may support the ATP Usage-related decisions. The process model is shown in Figure 9.

4.13 Process Models Master Planning and Factory

*Hubert Missbauer, Gerald Weigert, **Jesus Jimenez**, Lars Mönch, and Sebastian Knopp*

Event-driven process chains Master Planning – Factory

The master plan delivers quantities for the different facilities. Lots are generated based on the master plan. The lots are equipped with due dates. Release dates for the lots are either computed based on simple heuristics like backward and forward termination or based on optimization approaches. Intermediate due dates are derived. The proposed process model is shown in Figure 10.

4.14 Agents in the Level 3 Supply Chain

*Iris Lorscheid, **Scott Mason**, Irfan Ovacik, and Can Sun*

The "level 3" supply chain is defined by Infineon Semiconductor as a semiconductor company's supply chain. Within this level 3 supply chain exists a number of potential types of agents

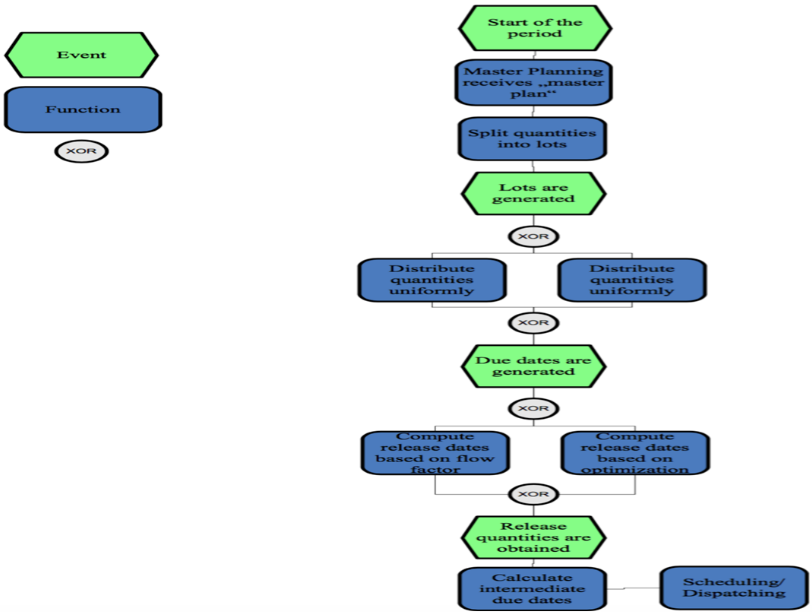

Figure 10 Process Model Master Planning – Factory.

(e.g., capacity planners, inventory planners, master planners, demand planners, supply chain planners, customer logistics managers, etc.). Regardless of its type, each agent has a set of attributes that include (but are not limited to) memory (i.e., the agent's knowledge base, rule set, and facts), the agent's role, and the coworkers with which the agent interacts. Further, each agent has abilities relating to individual strategies, input maintenance/analysis, and output analysis.

In terms of individual strategies, an agent can have specific strategy for dealing with inventory-related issues and ramp up/ramp down transitions, and another strategy for accommodating customer prioritization contingencies. As an example, consider a master planning agent (Figure 11; other agents could be inserted into this figure without loss of generality). During the input maintenance and analysis task, the agent may a) query databases and other data sources, b) ask the sales and marketing department for a demand input, and c) ask the factory planning department for a statement of available capacity. Once the master plan is complete, the output is analyzed for suitability. If the agent is satisfied, the results can be published to the production planning process (e.g., factory starts information) and/or the ATP process (e.g., a statement of supply). Otherwise, if the plan is not deemed to be suitable, the agent returns to the input maintenance and analysis task.

4.15 Top Down Reference Model

*Hans Ehm, Cathal Heavey, **Peter Lendermann**, Leon McGinnis, and Oliver Rose*

A High Level Semiconductor Supply Chain Model exists at a semiconductor company which can be considered a common denominator among all the stakeholders in Semiconductor

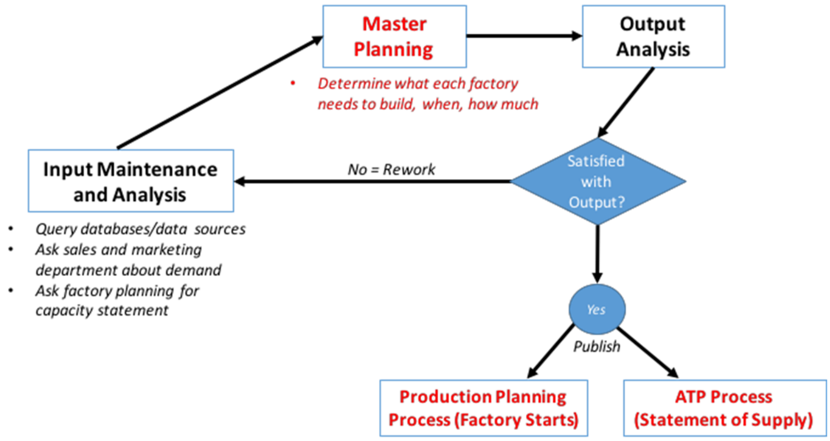

Figure 11 Process Model Master Planning – Factory.

Supply Chain Management and an appropriate starting point for the development of a Reference Model. However, business issues are typically identified on the lowest level process description level. A specific example would be an incident where it turned out that a change in shipment lot size for a particular product required different customs clearance procedure which was not taken into consideration during the Development and Master Planning stage for that product.

To avoid such incidents, the team agreed that a Reference Model could help if it also goes to the lowest level of process detail. However, for a process alignment on this level a common DSL (Domain Specific language) based on SYSML and containing PROS (Process, Role, Object, and State) would be needed. In a second stage this would also require an agreement between semiconductor manufacturers, software supplies, academia and other stakeholders to take full benefit of such a DSL in semiconductor supply chain. This could turn into an uphill task for the following reasons which need to be carefully taken care of:

- Low level process descriptions might contain some participants' IP which owners would want to protect. Once participants understand how a Top-Down Reference Model can be developed without disclosing company-sensitive information this restriction can be overcome as nobody wants to be left behind.
- Development of a Reference Model would require considerable resources which are only realistic if all participants can realise an ROI. After an initial publication and/or a panel discussion at WSC 2016, a European, US or Asia funded project could be the next step. The collaboration of International SEMATECH and JESSI, which ended up with the MIMAC models, could serve as a role model.

Participants

- Chen-Fu Chien
National Tsing Hua Univ., TW
- Stéphane Dauzère-Pérès
École des Mines de Saint-Etienne, FR
- Ton de Kok
TU Eindhoven, NL
- Hans Ehm
Infineon Technologies – München, DE
- Kenneth Fordyce
Arkieva – Wilmington, US
- José M. Framinán
University of Sevilla, ES
- Cathal Heavey
University of Limerick, IE
- Raphael Herding
FernUniversität in Hagen, DE
- Jesus Jimenez
Texas State University – San Marcos, US
- Adar Kalir
Intel Israel – Qiriat-Gat, IL
- Sebastian Knopp
École des Mines de Saint-Etienne, FR
- Peng-Chieh Lee
National Tsing Hua University – Hsinchu, TW
- Peter Lendermann
D-SIMLAB – Singapore, SG
- Iris Lorscheid
TU Hamburg-Harburg, DE
- Scott J. Mason
Clemson University, US
- Leon F. McGinnis
Georgia Inst. of Technology, US
- Hubert Missbauer
Universität Innsbruck, AT
- Lars Mönch
FernUniversität in Hagen, DE
- Irfan Ovacik
Intel Corporation – Chandler, US
- Thomas Ponsignon
Infineon Technologies – München, DE
- Oliver Rose
Universität der Bundeswehr – München, DE
- Can Sun
Infineon Technologies – München, DE
- Israel Tirkel
Ben Gurion University – Beer Sheva, IL
- Reha Uzsoy
North Carolina State Univ., US
- Gerald Weigert
TU Dresden, DE
- Jei-Zheng Wu
Soochow Univ. – Taiwan, TW

Report of Dagstuhl Seminar 16071

Pattern Avoidance and Genome Sorting

Edited by

Michael Albert[1], Miklós Bóna[2], István Miklós[3], and Einar Steingrímsson[4]

1 **University of Otago, NZ,** malbert@cs.otago.ac.nz
2 **University of Florida – Gainesville, US,** bona@ufl.edu
3 **Alfréd Rényi Institute of Mathematics – Budapest, HU,** Miklósi@ramet.elte.hu
4 **University of Strathclyde, GB,** einar@alum.mit.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16071 "Pattern Avoidance and Genome Sorting".

Seminar February 14–19, 2016 – http://www.dagstuhl.de/16071
1998 ACM Subject Classification D.3.1 Formal Definitions and Theory, Semantics, E.1 Data Structures, F.2 Analysis of Algorithms and Problem Complexity
Keywords and phrases evolutionary distance, lists, metrics, patterns, permutations, sorting
Digital Object Identifier 10.4230/DagRep.6.2.65
Edited in cooperation with Miklós Bóna

1 Executive Summary

Miklós Bóna

The seminar took place from February 14, 2016, to February 19, 2016. It had 36 participants, who were researchers in theoretical computer science, combinatorics, and molecular biology. It was a geographically diverse group, with participants coming from the US, Canada, Brazil, Germany, Iceland, the United Kingdom, Sweden, France, Slovakia, Hungary and New Zealand. The seminar featured 18 talks, three of which were hourlong talks, and an open problem session.

Numerous collaborative research efforts have been started. Here is a sampling.

Megan Martinez and Manda Riehl worked on a bijection between LP matchings (one of the RNA matchings described in Vincent Vatter's talk) and Klazar's nesting equivalent matchings. They studied a paper by Klazar and Aziza Jefferson's dissertation and made progress on the bijection.

István Miklós, Péter Erdős and Miklós Bóna worked on proving a log-convexity conjecture related to ordered degree sequences of bipartite graphs.

Brona Brejova and Manda Riehl discussed two potential future projects related to gene and species tree reconciliation. The most probable starting point is a project involving gene and species trees where a gene is allowed to duplicate a string inside itself. This situation was not allowed in previous models, however it seems that as long as the specific breakpoints are not reused from this insertion, a modification of the previous algorithms could still be effective.

Jay Pantone, David Bevan and Miklós Bóna collaborated on asymptotic enumeration of a balanced urns and balls model that was seen to be a step towards finding a better upper bound for a pattern avoidance enumeration problem.

We have all the reasons to believe that this, and many other joint research efforts that started during this seminar will lead to new results that would not have been possible without the seminar. Therefore, we strongly believe that the seminar was a success that we would like to repeat at some point in the future.

2 Table of Contents

3 Overview of Talks

3.1 Decomposition trees of permutations, and how to use them for a (realistic?) study of perfect sorting by reversals

Mathilde Bouvel (Universität Zürich, CH)

Joint work of Mathilde Bouvel, Cedric Chauve, Marni Mishna, Cyril Nicaud, Carine Pivoteau, and Dominique Rossin
Main reference M. Bouvel, M. Mishna, C. Nicaud, "Some simple varieties of trees arising in permutation analysis", in Proc. of the 25th Int'l Conf. on Formal Power Series and Algebraic Combinatorics (FPSAC 2013), DMTCS proceedings, 2013.
URL http://www.dmtcs.org/dmtcs-ojs/index.php/proceedings/issue/view/130
URL http://user.math.uzh.ch/bouvel/presentations/Poster2013_BMN.pdf

I will present the genome rearrangement problem of perfect sorting by reversals, and show its algorithmic solution by Bérard-Bergeron-Chauve-Paul. It uses the strong interval trees of permutations (whose definition will be recalled in the talk). Those trees are also known as (substitution) decomposition trees of permutations. I will present two results and a research project related to those trees.

First, I will show that the algorithm of Bérard-Bergeron-Chauve-Paul is polynomial on average (and with probability 1 as the size of the input goes to infinity). I will also describe average properties of commuting scenarios (a.k.a. separable permutations) for perfect sorting by reversals. These results are obtained using the tools of analytic combinatorics developed by Flajolet-Sedgewick.

Second, I will generalize these average properties (to some extent) to super-classes of the separable permutations. This demonstrates a phenomenon of convergence of a sequence of families of trees to the class of all permutations, whose analytic nature is essentially different. We only have a partial explanation of this phenomenon.

Finally, I will go back to the biological motivation, showing that all these models of trees do not represent well the strong interval trees obtained from the biological data. I will conclude by proposing a possibly better model, whose study is yet to be done.

3.2 Isometric Gene Tree Reconciliation

Brona Brejova (Comenius University in Bratislava, SK)

Joint work of Brona Brejova, Michal Sabo

The infinite sites model, introduced by Jian Ma et al (PNAS 2008), formalizes the problem of recovering the evolutionary history of a set of related genomes allowing a large set of evolutionary operations including insertions, deletions, duplications, and rearrangements. One of the steps of their polynomial-time algorithm reconciles a gene tree with a species tree under the assumption that both trees have exact branch lengths known. This assumption simplifies the problem compared to the typical reconciliation scenario without branch lengths. We show several mistakes in the original algorithm and provide a corrected and simplified version. We also discuss related open problems.

3.3 The SCJ small parsimony problem for weighted gene adjacencies

Cedric Chauve (Simon Fraser University – Burnaby, CA)

Joint work of Ashok Rajaraman, Yann Ponty, Cedric Chauve
Main reference A. Rajaraman, C. Chauve, Y. Ponty, "Assessing the Robustness of Parsimonious Predictions for Gene Neighborhoods from Reconciled Phylogenies: Supplementary Material", Bioinformatics Research and Applications, LNCS, Vol. 9096, pp. 260–271, Springer, 2015.
URL http://dx.doi.org/10.1007/978-3-319-19048-8_22

Reconstructing ancestral gene orders in a given phylogeny is a classical problem in comparative genomics. Most existing methods compare conserved features in extant genomes in the phylogeny to define potential ancestral gene adjacencies, and either try to reconstruct all ancestral genomes under a global evolutionary parsimony criterion, or, focusing on a single ancestral genome, use a scaffolding approach to select a subset of ancestral gene adjacencies, generally aiming at reducing the fragmentation of the reconstructed ancestral genome. We describe an exact algorithm for the small parsimony problem that combines both approaches. We consider that gene adjacencies at internal nodes of the species phylogeny are weighted, and we introduce an objective function defined as a convex combination of these weights and the evolutionary cost under the Single-Cut-or-Join (SCJ) model. The weights of ancestral gene adjacencies can e.g. be obtained through the recent availability of ancient DNA sequencing data, which provide a direct hint at the genome structure of the considered ancestor, or through probabilistic analysis of gene adjacencies evolution. The algorithm we propose is Fixed-Parameter Tractable (FPT) based on the dynamic programming algorithm by (Sankoff and Rousseau, 1975) and allows to sample co-optimal solutions.

3.4 Method of moments estimates for reversal and block transposition distances using symmetric group models

Niklas Eriksen (University of Örebro, SE)

Joint work of Niklas Eriksen, Axel Hultman
Main reference N. Eriksen, A. Hultman, "Expected reflection distance in G(r,1,n) after a fixed number of reflections", Annals of Combinatorics, 9(1):21–33, Springer, 2005.
URL http://dx.doi.org/10.1007/s00026-005-0238-y

The gene order of species change over time, and can hence be used to infer relationships between species. By estimating the number of reversals or block transpositions that separate two species, we get an idea of the time since speciation. Eriksen and Hultman introduced a permutation model with similar properties to the reversal model but much more tractable. In this model, method of moments estimates could be computed, corresponding to the reversal distance in a pure reversal model. We extend their results to obtain several estimates of the number of reversal and block transpositions that separate two species. We also give a method for estimating the proportion on reversals, which is a very important parameter.

3.5 Gene orders, median of permutations and related combinatorial problems

Sylvie Hamel (University of Montréal, CA)

Joint work of Sylvie Hamel and Robin Milosz
Main reference R. Milosz, S. Hamel, "Medians of Permutations: building constraints", Algorithms and Discrete Applied Mathematics, LNCS, Vol. 9602, pp. 264–276, Springer, 2016.
URL http://dx.doi.org/10.1007/978-3-319-29221-2_23

Our aim is to discuss the general problem of finding a consensus ranking, given a set of different rankings of a set of objects. Here, one assumes that a set of different rankings are proposed for a given set of strictly ordered elements, and one is looking for a ranking that is in closest agreement to all of these. Thus, the problem becomes that of finding the median of a set of permutations under a distance d. In part this is motivated by the classical gene order problem of comparative genomics, where the difference in the order of appearance of genes in the genome of different species is used to evaluate the evolutionary distances between them.

We give close attention to this median problem for the "Kendall-tau" distance, which corresponds to counting the number of order disagreements between pairs of elements of two permutations. The problem has been shown to be NP-complete for sets of m permutations, where $m > 3$, and the complexity is still unknown for $m = 3$. From an algorithmic point of view, we present a deterministic heuristic for this median problem, and derive some theoretical properties of the starting set of permutations that drastically reduce the search space for the medians of this set. In a more combinatoric point of view, we consider the interesting automedian cases (when a set of permutations is equal to the set of its medians), deriving some of its properties under group actions, shuffle operation, etc. Finally, we generalize this problem to the problem of "aggregating ranking with ties", with an application to the bioinformatic context of calculating such medians for biological data related to certain diseases.

3.6 Sorting with Forbidden Intermediates

Anthony Labarre (University Paris-Est – Marne-la-Vallée, FR)

Most genome rearrangement problems on permutations can be recast as constrained sorting problems, where the goal is to compute of a shortest sorting sequence of operations for a given permutation under the restriction that the set of allowed operations is fixed beforehand. However, biologists have known for more than a century that some of these mutations at a given point in time can be lethal to a given organism. We revisit those problems by adding a new constraint on the sequences to be computed: they must avoid a given set of forbidden intermediates, which correspond to species that cannot exist because the mutations that would be involved in their creation are lethal. We initiate this study by focusing on the case where the only mutations that can occur are algebraic transpositions, and give a polynomial time algorithm for solving that problem when the permutation to sort is an involution.

3.7 Efficient algorithms for permutation pattern matching

Marie-Louise Lackner (TU Wien, AT)

Joint work of Michael Albert, Marie-Louise Lackner, Martin Lackner and Vincent Vatter
Main reference M. H. Albert, M.-L. Lackner, M. Lackner, V. Vatter, "The complexity of pattern matching for 321-avoiding and skew-merged permutations", arXiv:1510.06051 [math.CO], 2015.
URL http://arxiv.org/abs/1510.06051v1

Given two permutations τ and π where π, the pattern, is shorter than τ, the Permutation Pattern Matching problem (PPM) asks whether π is contained in τ. In general, this problem is known to be NP-complete, implying that we may not hope for efficient algorithms to solve PPM. Two directions have been pursued so far in order to circumvent this hardness result: First, one can look for special cases in which PPM can be solved efficiently, i.e., in polynomial time. Second, one can try to find a parameter that explains the computational hardness of this problem and confine the combinatorial explosion to this parameter. In this talk, I will give an overview of the algorithms known so far and present one algorithm following the first approach and one taking the second one in more detail.

3.8 Social choice and permutation patterns

Martin Lackner (University of Oxford, GB)

Joint work of Martin Lackner, Marie-Louise Lackner
Main reference M. Lackner and M.-L. Lackner, "On the Likelihood of Single-Peaked Preferences", arXiv:1505.05852 [cs.GT], 2015.
URL http://arxiv.org/abs/1505.05852v1

In this talk I will discuss several combinatorial and algorithmic problems from Social Choice and how to they relate to questions about permutation patterns. In particular, some questions about structure in preferences can be answered by translating them to questions about permutation patterns. Also, I will present open problems about permutation patterns that arise from questions in Social Choice.

3.9 Counting and sampling genome rearrangement scenarios: a meeting-point of combinatorics and computer science

István Miklós (Alfréd Rényi Institute of Mathematics – Budapest, HU)

Joint work of István Miklós, Heather Smith
Main reference I. Miklós and H. Smith, "Sampling and counting genome rearrangement scenarios", BMC Bioinformatics, 16(Suppl 14):S6, Springer, 2015.
URL http://dx.doi.org/10.1186/1471-2105-16-S14-S6

Even for moderate size inputs, there are a tremendous number of optimal rearrangement scenarios, regardless what the model is and which specific question is to be answered. Therefore giving one optimal solution might be misleading and cannot be used for statistical inferring. Statistically well funded methods are necessary to sample uniformly from the solution space and then a small number of samples are sufficient for statistical inferring.

In this talk, we are going to give an overview of the state-of-the-art of sampling and counting rearrangement scenarios. The talk will focus on how combinatorial methods can be used in computational statistics.

3.10 The method of differential approximants

Jay Pantone (Dartmouth College – Hanover, US)

For decades, the method of differential approximants has been applied to the study of statistical mechanics to estimate the singularity structure of the generating function of a sequence of positive integers, using only a finite number of initial terms of the generating function. While all such approximations are of course only non-rigorous estimates, experience shows these estimates to be remarkably accurate.

Differential approximants can be extremely useful to combinatorialists. We provide several examples of combinatorial sequences for which no generating function is known or conjectured yet the method of differential approximants provides a very accurate approximation of the asymptotic behavior of the sequence. We then describe several extensions to the method that are in progress.

3.11 Parametric Analysis of an SCFG-based model for RNA structure prediction

Svetlana Poznanovikj (Clemson University, US)

The function of the RNA molecule is often dependent on its structure and so understanding how the RNA nucleotide chain folds onto itself is an important problem. Language-based methods for RNA structure prediction use stochastic context-free grammars (SCFGs). The SCFG developed by Knudsen and Hein is relatively simple and yet has been shown to achieve good accuracy compared to other grammars. We performed an analysis of the probability distribution induced by this grammar and in this talk I'll present some interesting properties that we found.

3.12 Longest increasing subsequences and log concavity

Bruce Sagan (Michigan State University – East Lansing, US), Miklós Bóna (University of Florida – Gainesville, US), and Marie-Louise Lackner (TU Wien, AT)

Joint work of Miklós Bóna, Marie-Louise Lackner, Bruce Sagan
Main reference M. Bóna, M.-L. Lackner, B. Sagan, "Longest increasing subsequences and log concavity", arXiv:1511.08653 [math.CO], 2015.
URL http://arxiv.org/abs/1511.08653

Let $\mathfrak{S}_n$ be the set of all permutations of $1, 2, \ldots, n$ viewed as sequences. Let $l_{n,k}$ be the number of $\pi \in \mathfrak{S}_n$ having a longest increasing subsequence of length k. This length is closely related to the Ulam distance between permutations which is used to model evolutionary distance in DNA research. William Chen conjectured that the sequence $l_{1,n}, l_{2,n}, \ldots, l_{n,n}$ is log concave which means that

$$l_{k-1,n} l_{k+1,n} \leq l_{n,k}^2$$

for all k. We also conjecture that if $i_{n,k}$ is the number of involutions in $\mathfrak{S}_n$ with longest increasing subsequence length k then $i_{1,n}, i_{2,n}, \ldots, i_{n,n}$ is log concave. We show that these two conjectures are strongly related. We also present evidence to support the truth of both. Our main tool is the Robinson-Schensted correspondence. Many other associated conjectures will be discussed.

3.13 Consequences of the no-coincidence assumption in comparative gene order

David Sankoff (University of Ottawa, CA)

The number of genes in plant and animal genomes tends to exceed 25,000, so that the coincidental gene orders must be very rare unless the genomes have inherited common orders. We discuss two consequences of this. One is the number of chromosomes inferred in an ancestral genome construction, especially in the context of ancient whole genome duplication, where inferring an incorrect number of chromosomes may require unlikely coincidences, The other is the inference of the locus of genome rearrangement on a phylogenetic tree. In this case only certain branches may be inferred to contain a breakpoint, otherwise coincidental changes must have happened. Both of these principles are key to recent reconstructions of ancestral gene orders.

3.14 An Update on Gene Family-free Genome Comparison

Jens Stoye (Universität Bielefeld, DE)

Joint work of Daniel Doerr, Louis Antonio Kowada, Jens Stoye, Simone Dantas
Main reference L. A. B. Kowada, D. Doerr, S. Dantas, J. Stoye, "New Genome Similarity Measures based on Conserved Gene Adjacencies", in Proc. of the 20th Annual Conf. on Research in Computational Molecular Biology (RECOMB'16), LNCS, Vol. 9649, pp. 204–224, Springer, 2016; pre-print available from author's webpage.
URL http://dx.doi.org/10.1007/978-3-319-31957-5_15
URL http://www.techfak.uni-bielefeld.de/~stoye/mypublications.html#KOW-DOE-DAN-STO-2016

Many methods in computational comparative genomics require gene family assignments as a prerequisite. While the biological concept of gene families is well established, their computational prediction remains unreliable. In this talk I will present recent results in a new line of research, in which family assignments are not presumed. We study several family-free approaches in detecting conserved structures, genome rearrangements and in reconstructing ancestral gene orders. This leads to a number of interesting combinatorial optimization problems, some of which are easily polynomial-time solvable, while others turn out to be surprisingly hard.

3.15 Positional Constraints for Rearrangements through Noncrossing Colored Partitions and Cycle Packings

Krister Swenson (University of Montpellier 2, FR)

Joint work of Krister Swenson, Mathieu Blanchette, Pijus Simonaitis
Main reference K. M. Swenson, M. Blanchette, "Models and algorithms for genome rearrangement with positional constraints", in Proc. of the 15th Int'l Workshop Algorithms in Bioinformatics (WABI'15), LNCS, Vol. 9289, pp. 243–256, Springer, 2015.
URL http://dx.doi.org/10.1007/978-3-662-48221-6_18

The number of moves is often the sole criterion used to measure the quality of a rearrangement scenario. A current challenge is to incorporate biological information into the gene-order evolutionary model in a manner that is computationally tractable. We present a model amenable towards positional constraints (as seen with chromatin conformation capture data), while elucidating connections between genome rearrangements and noncrossing colored partitions and cycle packings.

3.16 Breaking bad

Eric Tannier (University Claude Bernard – Lyon, FR)

The permutation as a model for gene order is flawed. It ignores the diversity of susceptibility to breakage across genomic regions, which is necessary even under a uniform random breakage model. I will propose a model of evolution of gene order by inversions where breakage probabilities vary across intergne regions and over time. It contains as a particular case the uniform breakage model on the nucleotidic sequence, where breakage probabilities are proportional to intergene region lengths. This is very different from the frequently used pseudo-uniform model where all intergene regions have the same probability to break. Estimations of rearrangement distances based on the pseudo-uniform model completely fail on simulations with the truly uniform model. I will propose new combinatorial and statistical problems with this model.

3.17 The substitution decomposition of RNA secondary structures

Vincent Vatter (University of Florida – Gainesville, US)

Joint work of Vincent Vatter and Aziza Jefferson
Main reference A. Jefferson, "The substitution decomposition of matchings and RNA secondary structures", Doctoral dissertation, University of Florida, 2015.
URL http://gradworks.umi.com/37/29/3729175.html

The substitution decomposition has proved to be a powerful tool for analyzing classes of permutations and of graphs (where it is known as the modular decomposition). I will discuss applications of the substitution decomposition to RNA secondary structures, which can be modeled by matchings.

3.18 On sequence segmentation problems

Tomáš Vinař (Comenius University in Bratislava, SK)

Joint work of Martina Višňovská, Tomáš Vinař, Broňa Brejová
Main reference M. Višňovská, T. Vinař, and B. Brejová, "DNA Sequence Segmentation Based on Local Similarity", in Information Technologies – Applications and Theory (ITAT), CEUR Workshop Proceedings, Vol. 1003, pp. 36–43, 2013.
URL http://ceur-ws.org/Vol-1003/

Most rearrangement models work with short segments (called markers, genes, atoms, etc.) that are considered atomic, i.e. they are long enough so that they are identifiable in a genomic sequence and at the same time they are not broken by rearrangement operations. It turns out that precomputing such segments is a difficult and interesting problem, and it is a major obstacle in applying algorithms for rearrangement analyses to real data. In some cases, the problem is in fact equivalent to reconstructing evolutionary history in a given model. In the talk I would discuss several approaches to solving this problem and also include some notes on an approach to analysis of segmental duplications that we have previously introduced.

Participants

- Michael Albert
University of Otago, NZ
- David Bevan
The Open University – Milton Keynes, GB
- Miklós Bóna
University of Florida – Gainesville, US
- Mathilde Bouvel
Universität Zürich, CH
- Marilia Braga
Inmetro – Duque de Caxias, BR
- Brona Brejova
Comenius University in Bratislava, SK
- Robert Brignall
The Open University – Milton Keynes, GB
- Cedric Chauve
Simon Fraser University – Burnaby, CA
- Anders Claesson
University of Strathclyde, GB
- Péter L. Erdös
Alfréd Rényi Institute of Mathematics – Budapest, HU
- Niklas Eriksen
University of Örebro, SE
- Pedro Feijão
Universität Bielefeld, DE
- Guillaume Fertin
University of Nantes, FR
- Sylvie Hamel
University of Montréal, CA
- Vít Jelínek
Charles University – Prague, CZ
- Anthony Labarre
University Paris-Est – Marne-la-Vallée, FR
- Marie-Louise Lackner
TU Wien, AT
- Martin Lackner
University of Oxford, GB
- Megan Martinez
Ithaca College, US
- István Miklós
Alfréd Rényi Institute of Mathematics – Budapest, HU
- Jay Pantone
Dartmouth College – Hanover, US
- Adeline Pierrot
University of Paris South XI, FR
- Yann Ponty
Ecole Polytechnique – Palaiseau, FR
- Svetlana Poznanovikj
Clemson University, US
- Manda Riehl
University of Wisconsin – Eau Claire, US
- Bruce Sagan
Michigan State University – East Lansing, US
- David Sankoff
University of Ottawa, CA
- Rebecca Smith
The College at Brockport, US
- Einar Steingrimsson
University of Strathclyde, GB
- Jens Stoye
Universität Bielefeld, DE
- Krister Swenson
University of Montpellier 2, FR
- Eric Tannier
University Claude Bernard – Lyon, FR
- Vincent Vatter
University of Florida – Gainesville, US
- Stéphane Vialette
University Paris-Est – Marne-la-Vallée, FR
- Tomáš Vinař
Comenius University in Bratislava, SK

Report from Dagstuhl Seminar 16072

Assessing Learning In Introductory Computer Science

Edited by

Michael E. Caspersen[1], Kathi Fisler[2], and Jan Vahrenhold[3]

1 **Aarhus University, DK,** mec@cs.au.dk
2 **Worcester Polytechnic Institute, US,** kfisler@cs.wpi.edu
3 **Universität Münster, DE,** jan.vahrenhold@uni-muenster.de

Abstract

This seminar discussed educational outcomes for first-year (university-level) computer science. We explored which outcomes were widely shared across both countries and individual universities, best practices for assessing outcomes, and research projects that would significantly advance assessment of learning in computer science. We considered both technical and professional outcomes (some narrow and some broad) as well as how to create assessments that focused on individual learners. Several concrete research projects took shape during the seminar and are being pursued by some participants.

Seminar February 14–19, 2016 – http://www.dagstuhl.de/16072
1998 ACM Subject Classification D.2 Software Engieering, D.3 Programming Languages, F.2 Analysis of Algorithms and Problem Complexity, F.3 Logics and Meanings of Programs, K.3.2 Computer and Information Science Education
Keywords and phrases Assessment, Learning Objectives
Digital Object Identifier 10.4230/DagRep.6.2.78
Edited in cooperation with Mirko Westermeier

1 Executive Summary

Jan Vahrenhold
Michael E. Caspersen
Kathi Fisler

The goal of the seminar was to focus on several broadly applicable learning outcomes for first year university computer science courses, looking at what it would take to understand and assess them in multiple pedagogic contexts.

In preparation for the seminar, we surveyed participants to get an understanding of a what could be a common denominator of CS1/2 learning outcomes, using the outcomes from the ACM CC 2013 curriculum as a starting point. We asked participants to (a) identify ones that are covered in their institution's CS1/2 courses, and (b) to identify ones that they have either experience or interest in investigating further. Participants also suggested objectives that were not included in CC 2013.

Of these candidate outcomes, we studied a subset during the seminar, as voted by the participants. We used breakout sessions to get small groups of participants to focus on individual outcomes, reporting on what is known about each outcome, its underlying challenges and/or relevant underlying theory, how to best assess it, and what sorts of research questions should be asked to advance educational research on that outcome. We had three separate sets of breakout sessions, so each participant had the chance to work on three outcomes in detail during the week. The discussion of some sessions was continued in a following session.

Rather than have most individual participants give talks, we ran three speed-dating poster sessions on the first afternoon: each person got to put up a poster on some outcome that they have studied, so others could see the research of other attendees.

In addition, we had three invited presentations focussing on workload and determinants of study success (Schulmeister), types of prior knowledge and their relation to study success (Theyssen), and Concept Inventories (Kaczmarczyk and Wolfman). The abstracts of these presentations are included in this report.

2 Table of Contents

3 Overview of Talks

3.1 A Concept Inventory Crash Course

Lisa C. Kaczmarczyk (San Diego, US) and Steven A. Wolfman (University of British Columbia – Vancouver, CA)

Concept Inventories (CIs) are short, low-stakes, often multiple-choice assessments used to guage students' learning surrounding specific concepts within and across courses. Their ease of administration–both having students write the CI and analysing results–makes them ideal as sustainable instruments for measuring the impact of changes in teaching, particularly pedagogical shifts. CIs have been widely used in science fields, most famously and earliest the physics Force Concept Inventory. Well-designed CIs aim to not only assess whether students have expert-like conceptions of course material but also to identify commonly occurring misconceptions through their questions' distractors. By providing evidence of difficulty learning fundamental concepts and of widespread misconceptions, CIs can be instruments of change in teaching practice. Designing a CI requires a careful, multi-step process to identify the often-narrow scope of the assessment, solicit common misconceptions among students, craft assessment items targeting these concepts and misconceptions, and validate the items with both experts and students. While progress has been made on CIs for some topics in computing, significant challenges remain including the diversity of programming languages, the rapid pace of change in the field, fundamentally different programming paradigms, and codification of processes such as design and debugging as concepts.

Partial List of Concept Inventory Publications of Interest to Computing Education Researchers

- W. K. Adams and C. E. Wieman. Development and validation of instruments to measure learning of expert-like thinking. Int'l. J. of Science Ed., 33(9):1289–1312, 2011.
- V. L. Almstrum, P. B. Henderson, V. Harvey, C. Heeren, W. Marion, C. Riedesel, L.-K. Soh, and A. E. Tew. Concept inventories in computer science for the topic discrete mathematics. In Working Group Reports on Innov. and Tech. in CS Ed., pages 132–145, 2006.
- Goldman, Ken, Gross, Paul, Heeren, Cinda, Herman, Geoffrey, Kaczmarczyk, Lisa, Loui, Michael, Zilles, Craig. (2010, June). "Setting the Scope of Concept Inventories for Introductory Computing Subjects". ACM Transactions on Computing Education (TOCE).
- Goldman, Ken, Gross, Paul, Heeren, Cinda, Herman, Geoffrey, Kaczmarczyk, Lisa, Loui, Michael, Zilles, Craig. (2008). "Identifying Important and Difficult Concepts in Introductory Computing Courses using a Delphi Process". Presented at the 39th Annual Technical Symposium on Computer Science Education (SIGCSE 2008), Portland, Oregon.
- R. R. Hake. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. American Journal of Physics, 66(1):64–74, 1998.
- Herman, Geoffrey L., Loui, Michael C., Kaczmarczyk, Lisa, Zilles, Craig. (2012, March).

"Describing the What and Why of Students' Difficulties in Boolean Logic". ACM Transactions on Computing Education (TOCE) 12(1), 3:1–28.
- Herman, Geoffrey, Kaczmarczyk, Lisa, Loui, Michael, Zilles, Craig. (2008). "Proof by Incomplete Enumeration and Other Logical Misconceptions". Presented at the Fourth International Computing Education Research Workshop (ICER 2008), Sydney, Australia.
- D. Hestenes, M. Wells, and G. Swackhamer. Force concept inventory. The Physics Teacher, 30(3):141–158, March 1992.
- Kaczmarczyk, Lisa, Petrick, Elizabeth, East, J. Philip, Herman, Geoffrey L. (2010). "Identifying Student Misconceptions of Programming". Presented at the 41st Annual Technical Symposium of Computer Science Education (SIGCSE 2010), Milwaukee, Wisconsin
- E. Mazur. Qualitative vs. quantitative thinking: are we teaching the right thing? International Newsletter on Physics Education, 32, April 1996.
- K. Karpierz and S. A. Wolfman. 2014. Misconceptions and concept inventory questions for binary search trees and hash tables. In Proceedings of the 45th ACM technical symposium on Computer science education (SIGCSE '14). ACM, New York, NY, USA, pp. 109–114.
- R. A. Streveler, R. L. Miller, A. I. Santiago-Roman, M. A. Nelson, M. R. Geist, B. M. Olds "Rigorous methodology for concept inventory development: Using the 'assessment triangle' to develop and test the thermal and transport science concept inventory (TTCI)", International Journal of Engineering Education Vol. 27, No. 5, pp. 968–984, 2011
- C. Taylor, D. Zingaro, L. Porter, K. C. Webb, C. B. Lee, and M. Clancy. Computer science concept inventories: past and future. Computer Science Education, 24(4):253–276, 2014

3.2 ZEITLast-Project (workload project): List of research activities with reference to computer science

Rolf Schulmeister (Universität Hamburg, DE)

The ZEITLast-project aimed to find out more about the real workload of our bachelor students by measuring the time needed to attend lectures and courses, for self study, preparing for exams, as well as used for extracurricular activities. Students used a web-based time budget or diary method to enter data daily during five months of a semester in order to cover the lecture period, the exam phase, and the self study activities during lecture free weeks. Data entered were controlled every day.

Time diary studies (time use, time budget) have been conducted in following IT curricula or computer science courses:

- B.Sc. Engineering Informatics, TU Ilmenau, no innovation, Summer 2010, 4th term: Workload 21.9 hours/week
- B.Sc. Engineering Informatics, TU Ilmenau, partly blocked modules, Summer 2011, 4th term: Workload 24.5 hours/week
- B.Sc. IT Security, St. Pölten, completely blocked modules, Winter 2011/12, 1st term: Workload 32 hours/week

- B.Sc. Informatik, Universität Paderborn, no innovation, Summer 2015, 2nd term: Workload 18.95 hours/week
- B.Sc. IT Security, St. Pölten, completely blocked modules, Winter 2015/16, 1st term: ongoing
- M.Sc. IT Security, St. Pölten, completely blocked modules, Winter 2015/16, 1st term: ongoing

The data of further 30 time diary studies in subjects other than computer science may be compared with the results of the computer science studies, e.g. architecture, mechanical engineering, culture, economy, education, mathematics, media science, media technology, mechatronics, ship building, medicine, electrical engineering, physics, etc. The mean workload in all 35 subjects tested so far is 23 hours per week. Even if the only data of the 14 weeks of lecturing are calculated, the mean workload is less than 30 hours per week and much less than expected by the Bologna system. An important finding is that the inter-individual variance and the intra-individual variance are enormously big, meaning that no student manages his or her time, attendance, and self study phases day for day or week for week in the same way. This finding throws a critical light at surveys that ask for data of one week and generalize them for a whole semester. Then mean values and percentages are misleading.

Some of the time use studies were followed up by a test on learning motivation and study habit based on an Integrated Model of Learning Activity and Motivation (according to Deci & Ryan, Julius Kuhl, and Thomas Martens). These tests were applied to students in education, economy, and computer science. The test generates motivational profiles of students focussing on learning behavior: anxiety, responsibility, distraction vs concentration, persistence vs procrastination, control of negative emotions etc.

Two meta-studies have been written in order to generalize the results of the empirical analyses:

1. “Auf der Suche nach Determinanten des Studienerfolgs”. in J. Brockmann/A. Pilniok (Hrsg.), Studieneingangsphase in der Rechtswissenschaft, Nomos: Baden-Baden 2014, S. 72–205.
   - 300 international studies dealing with determinants of learning success have been analyzed and summarized. This essay includes a paragraph on time diary analysis and an overview over international studies reaching comparable workload data (21–28 hours per week). The studies on study habits prove that traditional variables e.g. gender, social status, family etc. do not influence the learning outcome but motivational study habits, endurance, concentration, attendance etc. are determinants of GPA and learning success.
2. “Abwesenheit von Lehrveranstaltungen Ein nur scheinbar triviales Problem. Eine Meta-Studie von 300 empirischen Arbeiten.” Hamburg 2015.
   - An analysis of 300 international studies dealing with the attendance in lectures or absence of students from lectures and courses and the effect on the GPA. The degree of attendance is a good determinant of success in exams. Students missing three or more times a lecture or a seminar have lesser chances to reach an A or B.

References

1 Schulmeister, R. & Ch. Metzger (Hrsg.) (2011): Die Workload im Bachelor: Zeitbudget und Studierverhalten. Eine empirische Studie. Münster [u.a.]: Waxmann (report about the first 12 time use studies)

2 Schulmeister, Rolf, Christiane Metzger & Thomas Martens (2012): Heterogenität und Studienerfolg. Lehrmethoden für Lerner mit unterschiedlichem Lernverhalten. Paderborner Universitätsreden Heft 123. Paderborn.

3 Christiane Metzger, Rolf Schulmeister & Thomas Martens (2012): Motivation und Lehrorganisation als Elemente von Lernkultur. Zeitschrift für Hochschulentwicklung. ZFHE Jg.7 / Nr.3 (Juni 2012).
4 Rolf Schulmeister, Christiane Metzger: Zur Rolle der Lehrorganisation bei der Gestaltung des studentischen Selbststudiums In: Brigitte Kossek / Markus F. Peschl (Hg.): Digital Turn? Zum Einfluss digitaler Medien auf Wissensgenerierungsprozesse von Studierenden und Hochschullehrenden. V&R unipress Vienna University Press 2012, S. 77–92.

3.3 Types of prior knowledge and academic success in biology and physics

Heike Theyssen (Universität Duisburg-Essen, DE)

High dropout rates in science, mathematics and engineering programs of study give reason to investigate predictors for academic success. Academic success is influenced by a variety of general and subject-specific factors. So far, most studies in this field have focused on general, non-specific factors like grade point average (Dochy, Segers & Bühl, 1999). The project presented in this talk is part of the research unit "Academic learning and study success in the entry phase of science and technology study programs" (https://www.uni-due.de/zeb/alster/index_engl) funded by the German Research Foundation. In our project we focus on students' subject-specific prior knowledge that has proven to be a good predictor of learning success. According to Hailikari and her colleagues (2007, 2009, 2010) we distinguish different types of prior knowledge: knowledge of facts, knowledge of meaning, integration of knowledge and application of knowledge. In the talk, the theoretical background is introduced and the test instruments used to assess the different types of prior knowledge are presented.

4 Working groups

4.1 Decomposition and Algorithm Selection

Notes by Mirko Westermeier

The process of decomposition plays a central role while problem solving. The skills required to approach problems and split them into subtasks is essential not only for programmers but also for most (technical) problem solvers in life, regardless of their discipline. So it seems to be easy to motivate teaching these skills. Open questions are when to teach them, how they could optimally sequenced and whether decomposition should be taught explicitly.

4.2 Notional Machine

Notes by Mirko Westermeier

While trying to understand what hinders students by solving their programming tasks in a (for the teacher) straight-forward way, it sometimes turns out that their mental model of the machine on which the program runs is the problem. The biggest misconception here seems to be that the computer is somehow trying to understand what the student is doing.

The interesting part of this problem is, that most of these mental models ("Notional Machines") are consistent and therefore their behaviour is reasonable even if it can not be synchronized with reality when going into depth. The notional machine can be a collection of consistent misconceptions.

Attendees agreed that it is important to find out something about the teacher's own misconceptions and how they (if existent) match with the ones of the students. It seems to be important to know whether students have gaps in their mental models which lead to misconceptions and how tow design the teaching such that it is covering these gaps. The open research question here seems to be how to find and assess gaps in notional machines.

4.3 Processes and Decomposition

Notes by Mirko Westermeier

This breakout group discussed assessments in the context of understanding processes and decomposition of complex (programming) problems. An obvious starting point is debugging: while trying to fix a broken component or problem, students have to understand how the parts of the component work together and how the system changes over time processing different inputs. It seems that program comprehension and tracing of programs are prerequisites while becoming a successful programmer, but it is not so clear how to teach these skills. Attendees also were concerned that it could be dissatisfying to have students learn something that does not feel like coding to them. It could be necessary to do extra work to make these activities make sense to students. Therefore, teachers should have a clear awareness for these skills and need to motivate teaching them, which seems not be easy in an obvious way. It was mentioned that these activities need an amount of creativity and somehow collide with usual computer science matters. Attendees are unsure how to assess the understanding of decomposition strategies.

4.4 Social and Professional Issues

Notes by Mirko Westermeier

In this breakout group, attendees started with discussing what skills would be needed, required and useful in an industry position. It was stated that this is of course not the only important measure. Communities of Practice share these skills and a central question was what the intersections between common professional practices was. Another problem was the difference between these common professional practices and student's expectations and how to synchronize these: To eventually accomplish this and establish it in education, one not only needs to identify these professional practices but also to convey authenticity before even beginning to teach required skills.

Following questions would be to find out how to design curricula and tasks addressing identified communities of practice and how much of this would be able to teach and asses in student's first year.

4.5 Spatial Thinking

Notes by Mirko Westermeier

In this breakout session attendees discussed the connection of spatial thinking with computer science, its possibilities and what it implies. For example was asked whether spatial reasoning skills of different types of teachers were also different and if different sub-disciplines in computer science lead to different spatial reasoning. Considering students, a main touch point of spatial thinking with computer science is data structures, e.g. different expressions of binary, balanced trees and how the intuitive balancing model possibly translates into code. Also applied to mental models of heaps (graphical and text-based) the special manifestation of spatial thinking obviously makes a difference.

4.6 Tracing and Debugging

Notes by Mirko Westermeier

To define the skills neccessary for tracing and debugging a specific program, students have to develop a line-by-line understanding because the have to identify and leave out not-so-important parts. To follow the control flow and variable states, they have to use some kind of a notional machine and mentally execute parts of the program. Tracing therefore also requires a good understanding of the program's object graph, which object is doing what and how they change state during execution. This can be hard to accomplish depending on the complexity of components. In a linear program, students just have to follow the variables while in complex nested object interaction even using polymorphism, misconceptions in the mental model can become a real problem. Several tools to support better overall understanding were discussed. It was agreed that visualization tools (on different levels) help at first, but then a correct mental model is crucial.

The question how to assess tracing and debugging skills however remains hard. Attendees stated that it mostly never happened or failed because it was too hard. It seems important to watch students while trying to solve a problem and identify specific solving strategies in debugging. A defined set of these strategies should be testable. A belt system like in martial arts to state on which competence level a student is able to solve a task and wha she has to master before was discussed.

Most ideas from this breakout session were discussed later in the *Notional Machine* breakout.

4.7 Assessments with Learners in Mind

Steven A. Wolfman (University of British Columbia – Vancouver, CA)

The design of learning assessments – widely construed to include both research instruments and formative or summative instructional instruments – often focuses on qualities like validity and reliability that speak to properties of the data produced by the assessment. Both the process and results of assessment also have a powerful impact on learners' journey through their course of study. We believe that instructors and researchers will benefit from becoming more mindful of the impact of their assessments on the assessed learners' affinity for their course of study.

To change this situation, we propose a new dimension which – for the time being – we call affinity with which to design and consider assessments. Affinity is conceptually orthogonal to existing dimensions like validity and reliability. We anticipate that an assessment with high affinity empowers and motivates students to engage in their course of study and conversely, that an assessment with low affinity disengages and isolates students from their course of study. The affinity dimension will explicitly reflect the student voice in the assessment process, structured by established, research-based frameworks [Resnick and Shaffer, Ryan and Deci, Bandura, and Engle] exploring the fit and effect of instructional processes on students' mindset, motivation, and perceptions within a course of study.

We intend to explore, explain, and validate the nature of this affinity dimension and – in the process – develop a pair of instruments, one usable by researchers investigating this new affinity dimension to guide and frame the data they collect, and another related instrument usable by instructors to improve the affinity of their designed assessments. We hope that students will benefit from both of these through assessments with higher affinity that improve their learning process and trajectory.

This agenda will involve collecting and analysing a corpus of existing assessments embedded in their contexts; observing and interviewing students as they journey through an assessment process; observing and interviewing instructors and possibly researchers as they design and manage assessments; iteratively designing the instruments described above; and, of course, exploring the validity, reliability, and affinity of the assessment instruments and practice we design.

At Schloss Dagstuhl, over the course of three consecutive breakout sessions, we began this process by discussing the impact of assessments on students and their utility for students.

In search of concrete data to frame our conversation, we collected from the educators and researchers at the seminar a wide set of examples of assessments (assessments with no special expectation that they are exceptional) and selected a small set of exemplars of assessments (assessments where there is reason to believe they display best practices with regard to affinity). We brainstormed a set of questions representative of the student voice in the assessment process; and collected research-based frameworks to guide our analysis of the impact of assessment on students' mindset, motivation, and perceptions of their course of study. We are now engaged in iterative adaptation of an existing instrument for exploring users' journeys through use of a service to the context of assessment, using the student voice questions, research frameworks, and concrete examples described above.

Our next steps include two ongoing agendas and a set of precise goals. On an ongoing basis, we will continue to expand and organize our corpus of assessment examples. We will also begin the design of a "dashboard" learners can use to integrate the process and results of assessments into their ongoing course of study. This dashboard creates a frequent series of touchpoints between the learner and the instructor/researcher, which represent opportunities to both improve and explore the affinity of assessments.

5 Panel discussions

5.1 Assessment Techniques Brainstorming

Notes by Mirko Westermeier

To synchronize efforts across all participants, we discussed established assessment techniques and related open research questions.

Regarding the teacher's perspective, rapid diagnostics were mentioned to be an important tool to determine at what level instruction should start with beginners. For a continuous adjustments to student's abilities and progress, however, the focus should be on formative assessments. The question whether teachers are assessing what students actually had a chance to learn, directly relates to the concept of Constructive Alignment. The importance of measuring the process of student's work instead of just expecting the right answer was also highlighted: teachers should also look at intermediate steps and evaluate the process as such.

Most attendees agreed that a focus on the value of the assessment for the learner is of high importance while designing assessments. This could be achieved by using assessments and contexts which are relevant to the students' personal or professional life. Observational techniques, including time tracking, could be used to measure the students' workload, the dynamics of motivation, and – if applicable – the point of giving up.

Regarding programming assessment, there was a wide consensus that coding itself is only one of many important skills for students to learn. The ability to explain concepts to get to deep understanding as well as the ability to figure out simplest ways to test a program, an algorithm, or a component for correctness should also be considered a target for assessments.

Since developing soft skills like communication with colleagues and experts of different fields are very important for computer scientists and programmers, several attendents highlighted the importance of assessing these skills. Collaborative work, e.g., in programming, might be used to achieve this even in exams.

It was agreed upon that there is a high demand for validated assessment tools across all areas of computer science education research.

5.2 Next Steps Discussion

Notes by Mirko Westermeier

In the closing session, several topics regarding the seminar and research implications were discussed.

The attendees considered this first Dagstuhl seminar on computer science education research a most valuable addition to the current conferences, workshops, and meetings. Bringing people together and having time for focused work was appreciated as well as the informal meeting format which was used during the seminar several times. For example, some groups used a discussion format in which they concentrated on one researcher's work at a time. It was agreed upon that having mentoring workshops at all computer science education research venues would help to stay connected.

The attendees liked that the atmosphere in all breakout groups were very open and open to critique. To make it easier for newcomers to the computer science education research community to dive into it and get to know researchers with similar topic, Andrew J. Ko offered to write a document about the internal structure of the community and their body of knowledge (after the seminar, the participants provided feedback on the first draft which then was made available online; see `https://faculty.washington.edu/ajko/cer`).

The seminar was seen as a step towards formalizing research agendas in a way that they can be communicated to a broader audience. In addition to the focus of the current seminar, K–12 computer science education research should be considered in a similar setting.

Finally, several groups of attendees agreed about further collaborative work to continue discussion and work on the research questions that arose during this week.

6 Poster abstracts

6.1 Programming Education for Novices

Michael E. Caspersen (Aarhus University, DK)

I am interested in programming education for novices. Previously, my interest concentrated on higher education, but with the increased focus on computing and programming in secondary (and primary) education, it becomes relevant to rethink purpose, learning goals and didactics of programming education with these specific target groups in mind. In higher computing education, the perspective is education of to-be experts whereas in secondary (and even more so in primary) education, the purpose is computing, computational thinking and programming as general education and 'building'. In particular, the need for (pre- and in-service) training of teachers for primary and secondary school requires development of an associated computing and programming didactics.

6.2 KETTI – Competence Development of Student Teaching Assistants in Computer Science

Holger Danielsiek (Universität Münster, DE)

Joint work of Hubwieser, Peter; Krugel, Johannes; Magenheim, Johannes; Ohrndorf, Laura; Ossenschmidt, Daniel; Schaper, Niclas; Vahrenhold, Jan

The KETTI-project is a research cluster focused on developing a competence model for undergraduate teaching assistants in computer science along with corresponding instruments. As part of the deliverables of the project, training modules aligned with the desired competencies will be made available for academic use. As discussed in the introduction the focus of KETTI are rst-year courses in computer science. KETTI is funded by the BMBF and brings together researchers from computer science education, computer science, and psychology. The core of the project consists of four research groups at three large public German research universities (Technical University of Munich, Westfälische Wilhelms-Universität Münster, and University of Paderborn). These universities cover a wide range of student audiences. All departments have offered formal UTA training courses in the past.

To broaden the scope even further, KETTI includes six associated partners at five other German computer science department and two associated partners at two German mathematics departments. The associated partners at the computer science partners have not yet established formal UTA training courses. These partners are consulted on a regular basis to ensure that assumptions about student and UTA populations are not biased towards the departments with a tradition of UTA training. The partners at the mathematics departments are consulted to enable transfer of knowledge between disciplines and to better contrast general requirements and requirements specic to computer science.

6.3 A Method to Analyse Computer Science Students' Teamwork in Online Collaborative Learning Environments

Katrina Falkner (University of Adelaide, AU)

Joint work of Vivian, Rebecca; Falkner, Nickolas; Tarmazdi, Hamid

Although teamwork has been identified as an essential skill for Computer Science (CS) graduates, these skills are identified as lacking by industry employers, which suggests a need for more proactive measures to teach and assess teamwork. In one CS course, students worked in teams to create a wiki solution to problem-based questions. Through a case-study approach, we test a developed teamwork framework, using manual content analysis and sentiment analysis, to determine if the framework can provide insight into students' teamwork behavior and to determine if the wiki task encouraged students to collaborate, share knowledge, and self-adopt teamwork roles. Analysis revealed the identification of both active and cohesive teams, disengaged students, and particular roles and behaviors that were lacking. Furthermore, sentiment analysis revealed that teams moved through positive and negative emotions over the course of developing their solution, toward satisfaction. The findings demonstrate the value of the detailed analysis of online teamwork. However, we propose the need for automated measures that provide real-time feedback to assist educators in the fair and efficient assessment of teamwork. We present a prototype system and recommendations, based on our analysis, for automated teamwork analysis tools.

6.4 Plan Composition

Kathi Fisler (Worcester Polytechnic Institute, US)

Several decades ago, researchers identified "plan composition" as a programming-related task that students struggle to perform well. We have conducted a series of studies that revisit planning studies, most recently attempting to contrast how students using different programming languages (some functional, some procedural) approach similar problems. We discuss our efforts to modernize planning studies to contexts framed by lightweight scripting for data analysis. Our results show the role that libraries and built-ins play in guiding how students decompose problems, and identify potential student misconceptions about the costs associated with using built-in operations. Overall, our goal is to help reframe research questions about planning for today's programming contexts.

6.5 Critiquing CS Assessment from a CS for All Lens

Mark Guzdial (Georgia Institute of Technology – Atlanta, US)

Not everyone who learns CS is going to want to be a software engineer. Then why teach them CS? Perhaps there are reasons to learn CS that are not about software development. As computing pervades many professions, there are reasons to learn to program that are relevant to those professions, those communities of practice, that may not be relevant to software engineers. For example, computational scientists and engineers may not care about developing code to professional standards, since their programs may not be meant to be used beyond a single time. If we have different learning outcomes for different communities of practice, assessment has to change, too. I argue that we have to consider what the learner wants to do and wants to be (i.e., their desired Community of Practice) when assessing learning. Different CoP, different outcomes, different assessments.

6.6 Communicating through Sketches and Diagrams

Geoffrey L. Herman (University of Illinois – Urbana Champaign, US)

Communicating through sketches and diagrams is a foundational skill in students' problem solving, yet little is known about how or why students use sketches and or how they learn about them. We are exploring students' representational fluency by exploring the similarities and distinctions between experts and novices in their use and production of sketches during problem solving. We are conducting clinical interviews in two domains: digital logic and trusses. While digital logic sketches are visually distinct from real-world implementations, trusses rely on fascimiles of real-world objects. We will share preliminary findings about students' use of sketching in digital logic contexts.

6.7 Neo-Piagetian Theory and the Novice programmer

Raymond Lister (University of Technology – Sydney, AU)

Piagetian-based cognitive development theories (Flavell, 1977; Morra, Gobbo, Marini, & Sheese, 2007; Piaget, 1952) provide a framework for describing the domain-specific development of cognition. In the case of programming, students exhibit characteristics at each of the neo-Piagetian sensorimotor, preoperational, and concrete operational stages. At the least mature stage of cognitive development, the sensorimotor stage, a novice programmer has difficulty tracing (i.e. manually executing) code. These students tend to see code as an ad hoc collection of tricks. At the next more mature stage, a preoperational novice has a more systematic grasp of programming, and can trace code with some accuracy. However, a preoperational programmer tends not to abstract from the code itself. Instead, the preoperational programmer favours inductive inference, That is, the preoperational programmer tends to infer what a piece of code does from input/output pairs, which are found by tracing the code. It is only at the third stage, the concrete operational stage, that the novice programmer begins to reason about abstractions of code, simply by reading the code. A defining characteristic of the concrete operational stage is the ability to reason about the concepts of conservation, reversibility and transitive inference. Neo-Piagetian theory emphasises that developmentally appropriate exposure to the domain of knowledge is paramount to the progression between developmental stages of reasoning. Unfortunately, contemporary teaching methods for programming assume that students begin at the concrete operational level.

6.8 Educational Data Mining

Andreas Mühling (TU München, DE)

Educational Data Mining allows to search for pattern in large amounts of data that are not directly observable. As such, the methods are often exploratory in nature and therefore well suited for the context of CSEd where an accepted body of fundamental results concerning the teaching and learning of the subject is still not avaiable. A method of aggregating concept map data has been used to investigate the effects of CS education on the structural configuration of knowledge in learners. The results show that artifacts of a curriculum can be traced to the knowledge structures of learners. Also, the large scale Bebras contest data has been used investigate the underlying psyhcometric structure of the items. Here, the results indicate that success inthe contest is determined to a large extent by visuo-spatial thinking.

6.9 Teaching and Learning a First Programming Language

Anthony Robins (University of Otago, NZ)

I am interested in novices learning a first programming language, in particular why they fail and why they succeed in a first programming course (CS1). I believe that the well explored distinction between novice and expert programmers is less important than the distinction between ineffective and effective novices. I argue that the early stages of learning in CS1 are crucial. Initial success or failure builds rapidly on itself to create momentum towards a final successful or unsuccessful outcome (the"learning-edge momentum" effect described in Robins, 2010).

It is particularly important to understand the process of novice learning as programming and related computational thinking topics are increasingly being introduced into the school curriculum, even elementary/primary school, in several countries around the world.

6.10 Novice Programmers' Difficulties with Program Dynamics and Misconceptions of the so-called Notional Machine

Juha Sorva (Aalto University, FI)

In this poster, I present an overview of some of the main themes of my research over the past five years. These themes include: novice programmers' difficulties with program dynamics and misconceptions of the so-called notional machine; interactive program visualization for education; the research-based instructional design of interactive electronic textbooks for computing education, and the benchmarking of students conceptual knowledge and programming skill across institutions.

6.11 Qualitative Feedback of Program Code and Programs

Martijn Stegeman (University of Amsterdam, NL)

Martijn Stegeman is finalising plans for a research project on the use of feedback in programming courses. During these courses, it is common to provide qualitative feedback on the program code as well as on the resulting programs, while students (hopefully) use this feedback to better understand quality norms and improve on new programs they write. Interestingly, the norms used by teachers are usually very personal and subjective, although there appears to be a lot of commonality. This project aims to produce understanding of the norms in use, and the way feedback is integrated into programming courses.

6.12 Potential Factors Indicating Success in an Algorithms and Data Structures Course

Jan Vahrenhold (Universität Münster, DE)

Joint work of Danielsiek, Holger

We report on first steps towards identifying factors indicating students' performance in an Algorithms and Data Structure course. We discuss a study undertaken to investigate the predictive and explanation power as well as the limits of weekly test items based on concept inventory questions, homework grades, and performance in a preceding CS1 course. We relate our findings for two subgroups to results on academic success in general and performance in a CS1 course in particular.

6.13 (Mis)conceptions Surrounding Exponential Growth in the Foundations of Computing Concept Inventory

Steven A. Wolfman (University of British Columbia – Vancouver, CA)

Exponential growth commonly appears as a topic in foundations of computing textbooks and exams. In our interviews with foundations of computing course instructors, about half mentioned exponential growth as an important but tricky topic for students. In our think-aloud interviews with students working problems not directly related to exponential growth, students nonetheless often expressed confusion around or simply misused the concept. As a result, we gathered and analyzed data on students' definitions of "exponential growth" from an open-ended question on a low-stakes assessment. Then, we designed two closed-ended (multiple choice) concept inventory questions on exponential growth based on this data, instructors' and students' comments in interviews, and our own experience. We performed think-aloud interviews with students working these closed-ended questions and administered the closed-ended questions to several cohorts of students (and several hundred students) in our three required foundations of computing courses, adapting the questions in response to strengths and weaknesses in the distractors exposed by both think-aloud and simple multiple-choice responses. The data show that students have persistent and distressing misconceptions around the concept of exponential growth, with the three most resilient distractors showing a graph curving "up and to the right", a function described as growing "more and more quickly", and the simple quadratic function n^2. Further research will be needed to connect these misconceptions to any impact on students' ability to judge and select appropriate approaches to computational problems.

Participants

- Michael E. Caspersen
Aarhus University, DK
- Holger Danielsiek
Universität Münster, DE
- Brian Dorn
University of Nebraska, US
- Katrina Falkner
University of Adelaide, AU
- Sally Fincher
University of Kent, GB
- Kathi Fisler
Worcester Polytechnic Inst., US
- Mark Guzdial
Georgia Institute of Technology – Atlanta, US
- Geoffrey L. Herman
University of Illinois – Urbana Champaign, US
- Lisa C. Kaczmarczyk
San Diego, US
- Andrew J. Ko
University of Washington – Seattle, US
- Michael Kölling
University of Kent, GB
- Shriram Krishnamurthi
Brown Univ. – Providence, US
- Raymond Lister
University of Technology – Sydney, AU
- Briana Morrison
Georgia Institute of Technology – Atlanta, US
- Jan Erik Moström
University of Umeå, SE
- Andreas Mühling
TU München, DE
- Anthony Robins
University of Otago, NZ
- Rolf Schulmeister
Universität Hamburg, DE
- Carsten Schulte
FU Berlin, DE
- R. Benjamin Shapiro
Univ. of Colorado – Boulder, US
- Beth Simon
University of California – San Diego, US
- Juha Sorva
Aalto University, FI
- Martijn Stegeman
University of Amsterdam, NL
- Heike Theyssen
Universität Duisburg-Essen, DE
- Jan Vahrenhold
Universität Münster, DE
- Mirko Westermeier
Universität Münster, DE
- Steven A. Wolfman
University of British Columbia – Vancouver, CA

Report from Dagstuhl Seminar 16081

Scheduling

Edited by

Nikhil Bansal[1], Nicole Megow[2], and Clifford Stein[3]

1 **TU Eindhoven, NL,** n.bansal@tue.nl
2 **TU München, DE,** nmegow@ma.tum.de
3 **Columbia University, USA,** cliff@ieor.columbia.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16081 "Scheduling". The seminar was centered around recent new developments, discussion of open problems and exploring future research directions within the broader scheduling community.

Seminar February 21–26, 2016 – http://www.dagstuhl.de/16081
1998 ACM Subject Classification C.4 Performance of Systems
Keywords and phrases approximation algorithms, scheduling, optimization
Digital Object Identifier 10.4230/DagRep.6.2.97
Edited in cooperation with Bouke Cloostermans

1 Executive Summary

Nikhil Bansal
Nicole Megow
Clifford Stein

This fourth meeting in a series of Dagstuhl "Scheduling" seminars had two major objectives. Firstly, it offered a forum for presenting recent scheduling results of high impact and new techniques which may be useful for solving important and long-standing open problems. The second major objective was to debate and explore future research directions, discuss important open problems, and foster new collaborations with a particular attention to interactions with application areas, both in academia and industry.

The organization of the meeting differed from the previous Dagstuhl "Scheduling" seminars by not inviting a different community to interact. Despite (or perhaps because of) the success of the cross-discipline events, there was an explicit desire to dedicate a seminar explicitly to recent advances and new research trends within the algorithmics/math programming scheduling community. This setting allowed for very high technical level talks and deep discussions on recent scheduling results, new techniques, and discussions on important open problems. The program included 15 invited main talks, 10 short spot-light talks, open problem sessions in the beginning of the week, and ample unstructured time for research and interaction. The overall atmosphere among the 45 participants was very interactive and oriented towards solving problems (also initiated by the few well-chosen application-driven talks) within new collaborations.

2 Table of Contents

3 Overview of Talks

3.1 Online Scheduling: New and Old Problems

Susanne Albers (TU München, DE)

We study problems in online scheduling, where jobs arrive incrementally over time and sequencing decisions must be made without knowledge of any future input. In the first part of the talk we investigate energy-efficient scheduling, where jobs are processed by variable-speed processors with the objective to minimize energy consumption. We focus on problems where jobs have deadlines and present results for multi-processor environments. Platforms with homogeneous and heterogeneous processors are considered. In the second part of the talk we revisit classical online makespan minimization. We review and develop results for various advanced problem settings in which online strategies are given extra power or information to process the incoming job sequence. In the scenarios addressed in the talk an algorithm (a) has a reordering buffer, (b) may migrate jobs, (c) is allowed to construct several candidate schedules, or (d) has some information about the job sequence.

3.2 On Speed Scaling with a Sleep State

Antonios Antoniadis (MPI für Informatik – Saarbrücken, DE)

Joint work of Susanne Albers, Antonios Antoniadis, Chien-Chung Huang and Sebastian Ott
Main reference A. Antoniadis, C. Huang, S. Ott, "A Fully Polynomial-Time Approximation Scheme for Speed Scaling with Sleep State", in Proc. of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'15), pp. 1102–1113, SIAM, 2015; pre-print available from author's webpage.
URL http://dx.doi.org/10.1137/1.9781611973730.74
URL http://www.cse.chalmers.se/~huangch/work/energysleep.pdf

We consider classical deadline-based preemptive scheduling of jobs in a computing environment equipped with both dynamic speed scaling and sleep state capabilities: Each job is specified by a release time, a deadline and a processing volume, and has to be scheduled on a single, speed-scalable processor that is supplied with a sleep state. In the sleep state, the processor consumes no energy, but a constant wake-up cost is required to transition back to the active state. In contrast to speed scaling alone, the addition of a sleep state makes it sometimes beneficial to accelerate the processing of jobs in order to transition the processor to the sleep state for longer amounts of time and incur further energy savings. The goal is to output a feasible schedule that minimizes the energy consumption.

We shortly exhibit the proof that the optimization problem is NP-hard and present a fully polynomial-time approximation scheme for the problem, which is based on a transformation to a non-preemptive variant of the problem. We conclude by discussing challenges that arise when trying to extend existing algorithms for the problem to the multiprocessor case.

3.3 Packing Small Vectors

Yossi Azar (Tel Aviv University, IL)

Joint work of Yossi Azar, Ilan Reuven Cohen, Amos Fiat, Alan Roytman
Main reference Y. Azar, I. R. Cohen, A. Fiat, A. Roytman, "Packing Small Vectors", in Proc. of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'16), pp. 1511–1525, SIAM, 2016; pre-print available from author's webpage.
URL http://dx.doi.org/10.1137/1.9781611974331.ch103
URL http://cs.tau.ac.il/~alanr/publications/FracOnVBinPack.pdf

Online d-dimensional vector packing models many settings such as minimizing resources in data centers where jobs have multiple resource requirements (CPU, Memory, etc.). However, no online d-dimensional vector packing algorithm can achieve a competitive ratio better than d. Fortunately, in many natural applications, vectors are relatively small, and thus the lower bound does not hold. For sufficiently small vectors, an $O(\log d)$-competitive algorithm was known. We improve this to a constant competitive ratio, arbitrarily close to $e \approx 2.718$, given that vectors are sufficiently small. We give improved results for the two dimensional case. For arbitrarily small vectors, the First Fit algorithm for two dimensional vector packing is no better than 2-competitive. We present a natural family of First Fit variants, and for optimized parameters get a competitive ratio ≈ 1.48 for sufficiently small vectors. We improve upon the 1.48 competitive ratio – not via a First Fit variant – and give a competitive ratio arbitrarily close to 4/3 for packing small, two dimensional vectors. We show that no algorithm can achieve better than a 4/3 competitive ratio for two dimensional vectors, even if one allows the algorithm to split vectors among arbitrarily many bins.

3.4 Applications of scheduling theory to realtime computing

Sanjoy K. Baruah (University of North Carolina at Chapel Hill, US)

Scheduling theory is one of the foundational cornerstones of the discipline of real-time computing. Prior collaborative research between the real-time computing and scheduling theory communities has yielded some important and interesting results; I will attempt to identify additional open problems in real-time computing that may be of interest to researchers in scheduling theory, and that are perhaps best solved via collaboration between the communities.

3.5 The Primal-Dual Approach to Online Optimization Problems

Niv Buchbinder (Tel Aviv University, IL)

The primal-dual method is one of the fundamental design methodologies in the areas of linear programming, combinatorial optimization, and approximation algorithms. In the area of online algorithms the primal-dual method emerged as a unifying framework to the design and analysis of online algorithms. The method has been shown to be applicable to many central online problems such as paging, routing, scheduling, online set cover, and graph connectivity problems.

In this talk I will provide a short introduction to the primal-dual method for online combinatorial optimization. I will then discuss several recent extensions of the method to online learning and convex objective functions, and will also present several open questions.

3.6 Probabilistic RT scheduling

Liliana Cucu-Grosjean (INRIA – Paris, FR)

Probabilistic real-time scheduling considers systems that have some parameters described by probability distribution. As expected this makes the problem harder while not any probabilistic description is appropriate for calculating the response time.

3.7 A Brief History of Speedup Factors for Uniprocessor EDF and Fixed Priority Scheduling

Robert Davis (University of York, GB)

The performance of real-time scheduling algorithms can be compared using the resource augmentation bound or speedup factor. The Speedup Factor comparing two real-time scheduling algorithms X and Y is given by the minimum factor by which the speed of the processor needs to be increased so as to ensure that any task set that is schedulable under algorithm Y is guaranteed to also be schedulable under algorithm X. In this talk we give a brief tour of the exact speedup factors which have been derived for comparisons between pre-emptive and non-pre-emptive, Earliest Deadline First and Fixed Priority scheduling (EDF-P, EDF-NP, FP-P, and FP-NP) on a single processor. The scope of the problem is uniprocessor systems, assuming the sporadic (real-time) task model, and considering three different classes of task set (with implicit, constrained and arbitrary deadlines).

We begin with results for FP-P v EDF-P published in 2009 and results for FP-NP v. EDF-NP published in 2010. In both cases, recent work published in 2015 has closed the gap between upper and lower bounds providing exact speedup factors, finally closing these problems for all three classes of task set. Further, we note that all of the exact speedup factors for FP-P v EDF-P and FP-NP v. EDF-NP continue to hold when simple linear (sufficient) schedulability tests are used for FP, along with Deadline Monotonic priority assignment (even when it is not optimal). These latter results are under review.

Recent work comparing the non-pre-emptive and pre-emptive paradigms has resulted in exact speedup factors for EDF-NP v EDF-P, FP-NP v. EDF-P, and FP-NP v. FP-P. (Since non-pre-emptive scheduling suffers from the so called long task problem, these speedup factors are parametric in C_{max}/D_{min}, the ratio of the longest execution time to the shortest deadline). To the best of our knowledge, proving the exact speedup factor for FP-P v FP-NP remains an open problem and is discussed in another talk.

3.8 Wireless Scheduling

Magnus M. Halldórsson (Reykjavik University, IS)

We examine scheduling in wireless networking, focusing on throughput and latency in link scheduling at the MAC layer in physical (or SINR) models of interference. These can be captured as packing and partitioning problems in independence systems with complicated feasibility predicate. While throughput (or Capacity) has been well solved in most contexts, much less has been achieved for latency (or Link Scheduling). We discuss recent approach based on abstracting the independence system as a graph, show that it achieves improved approximations, but that it also has inherent limitations. We also touch on the challenging question of how best to model real wireless environments with algorithmically usable models.

3.9 Fair Scheduling via Iterative Quasi-Uniform Sampling

Sungjin Im (University of California – Merced, US)

Joint work of Benjamin Moseley

We consider minimizing the ℓ_k-norms of flow time on a single machine offline using a preemptive scheduler for $k \geq 1$. We show the first constant approximation for the problem, improving upon the previous best $O(\log \log P)$-approximation by Bansal and Pruhs (FOCS 09 and SICOMP 14) where P is the ratio of the maximum job size to the minimum. Our main technical ingredient is a novel combination of quasi-uniform sampling and iterative rounding, which is of interest in its own right.

3.10 Approximation Schemes for Machine Scheduling with Resource (in-)dependent Processing Times

Klaus Jansen (Universität Kiel, DE)

Joint work of Klaus Jansen, Marten Maack and Malin Rau
Main reference K. Jansen, M. Maack, M. Rau, "Approximation schemes for machine scheduling with resource (in-)dependent processing times", in Proc. of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'16), pp. 1526–1542, SIAM, 2016.
URL http://dx.doi.org/10.1137/1.9781611974331.ch104

We consider two related scheduling problems: resource constrained scheduling on identical parallel machines and a generalization with resource dependent processing times. In both problems, jobs require a certain amount of an additional resource and have to be scheduled minimizing the makespan, while at every point in time a given resource capacity is not exceeded. We present a method to obtain asymptotic fully polynomial approximation schemes (AFPTAS) for the problems.

3.11 On the strength of the Configuration-LP for the Maximum Budgeted Allocation Problem

Christos Kalaitzis (EPFL Lausanne, CH)

Joint work of Aleksander Madry, Alantha Newman, Lukáš Poláček and Ola Svensson
Main reference C. Kalaitzis, A. Madry, A. Newman, L. Polácek, O. Svensson, "On the configuration LP for maximum budgeted allocation", Math. Progr., 154(1–2):427–462, Springer, 2015.
URL http://dx.doi.org/10.1007/s10107-015-0928-8

The MBA problem is the problem of assigning items to budget-constrained agents (each of which is willing to pay up to a certain amount for each item), such that the budgets of the agents are respected and the total amount the players pay is maximized. In this talk, I will give a few insights as to why the Configuration-LP is stronger than the natural LP-relaxation for the problem, a fact which also leads to the best approximation ratio we know for the problem.

3.12 The department chair's scheduling problem

Samir Khuller (University of Maryland – College Park, US)

The objective function is simple. I'd like to schedule 1hr meetings with release times and deadlines and minimize the number of days I have to go into campus for. We called in "Active time minimization".

3.13 Practice Driven Scheduling Models

Retsef Levi (MIT – Camridge, US)

In this overview talk, we would discuss several practical examples of scheduling decision in various application domains, such as maintenance, inventory management, warehousing, and hospital operations. The practical examples are chosen to illustrate some potentially new theoretical models and algorithmic challenges with the hope to inspire new directions in the academic work.

3.14 The Lasserre/Sum-of-Squares hierarchy for friends

Monaldo Mastrolilli (IDSIA – Manno, CH)

The Lasserre/Sum-of-Squares hierarchy is a systematic procedure to strengthen LP relaxations by constructing a sequence of increasingly tight formulations. For a wide variety of optimization problems, this approach captures the convex relaxations used in the best available approximation algorithms.

The goal of this tutorial is to introduce the non-expert to this technique by giving the essential intuition behind the definition and key properties. We discuss some simple examples, applications and limitations.

3.15 Scheduling Parallel DAG Jobs Online

Benjamin J. Moseley (Washington University – St. Louis, US)

Joint work of Kunal Agrawal, Jing Li, Kefu Lu
Main reference K. Agrawal, J. Li, K. Lu, B. Moseley, "Scheduling Parallel DAG Jobs Online to Minimize Average Flow Time", in Proc. of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'16), pp. 176–189, SIAM, 2016.
URL http://dx.doi.org/10.1137/1.9781611974331.ch14

In this talk, we discuss the problem of scheduling parallelizable jobs online. Each parallel job is modeled as a DAG where each node is a sequential task and each edge represents dependence between tasks. Previous work online has focused on a model of parallelizability known as the arbitrary speed-up curves setting. However, the DAG model is more widely used by practitioners, since many jobs generated from parallel programming languages and libraries can be represented in this model. Little is known for this model in the online setting with multiple jobs. The DAG model and the speed-up curve models are incomparable and algorithmic results from one do not imply results for the other.

Previous work has left open the question of whether an online algorithm can be $O(1)$-competitive with $O(1)$-speed for average flow time and maximum flow time in the DAG setting. In this talk will give the first scalable algorithms which are $(1+\epsilon)$-speed $O(1)$-competitive for average flow time and maximum flow time for any $\epsilon > 0$.

3.16 A Lasserre-based $(1+\epsilon)$-approximation for Makespan Scheduling with Precedence Constraints

Thomas Rothvoss (University of Washington – Seattle, US)

Joint work of Elaine Levey
Main reference E. Levey, T. Rothvoss, "A Lasserre-based $(1+\epsilon)$-approximation for Pm|p$_j$ = 1; prec |Cmax", arXiv:1509.07808 [cs.DS], 2015.
URL http://arxiv.org/abs/1509.07808v1

In a classical problem in scheduling, one has n unit size jobs with a precedence order and the goal is to find a schedule of those jobs on m identical machines as to minimize the makespan. It is one of the remaining four open problems from the book of Garey & Johnson whether or not this problem is NP-hard for $m = 3$.

We prove that for any fixed epsilon and m, a Sherali-Adams/Lasserre lift of the time-index LP with slightly super poly-logarithmic number of rounds provides a $(1+\epsilon)$-approximation.

The previously best approximation algorithms guarantee a $2 - 7/(3m+1)$-approximation in polynomial time for $m >= 4$ and 4/3 for $m = 3$. Our algorithm is based on a recursive scheduling approach where in each step we reduce the correlation in form of long chains.

Our method adds to the rather short list of examples where hierarchies are actually useful to obtain better approximation algorithms.

3.17 Online Algorithms for Multi-Level Aggregation

Jiri Sgall (Charles University – Prague, CZ)

Joint work of Marcin Bienkowski, Martin Böhm, Jaroslaw Byrka, Marek Chrobak, Christoph Dürr, Lukáš Folwarczný, Lukasz Jez, Nguyen Kim Thang, and Pavel Veselý
Main reference M. Bienkowski, M. Böhm, J. Byrka, M. Chrobak, C. Dürr, L. Folwarczny, L. Jez, J. Sgall, K. T. Nguyen, P. Veselý, "Online Algorithms for Multi-Level Aggregation", arXiv:1507.02378 [cs.DS], 2015.
URL http://arxiv.org/abs/1507.02378v2

In the Multi-Level Aggregation Problem (MLAP), requests arrive at the nodes of an edge-weighted tree T, and have to be served eventually. A service is defined as a subtree X of T that contains its root. This subtree X serves all requests that are pending in the nodes of X, and the cost of this service is equal to the total weight of X. Each request also incurs waiting cost between its arrival and service times. The objective is to minimize the total waiting cost of all requests plus the total cost of all service subtrees.

MLAP is a generalization of some well-studied optimization problems; for example, for trees of depth 1, MLAP is equivalent to the TCP Acknowledgment Problem, while for trees of depth 2, it is equivalent to the Joint Replenishment Problem. Aggregation problem for trees of arbitrary depth arise in multicasting, sensor networks, communication in organization hierarchies, and in supply-chain management. The instances of MLAP associated with these applications are naturally online, in the sense that aggregation decisions need to be made without information about future requests.

Constant-competitive online algorithms are known for MLAP with one or two levels. However, it has been open whether there exist constant competitive online algorithms for trees of depth more than 2. Addressing this open problem, we give the first constant competitive online algorithm for networks of arbitrary (fixed) number of levels. The competitive ratio is $O(D^4 2^D)$, where D is the depth of T. The algorithm works for arbitrary waiting cost functions, including the variant with deadlines.

3.18 Bounded Preemptions in Real-time Scheduling

Hadas Shachnai (Technion – Haifa, IL)

Scheduling problems are traditionally classified as either preemptive, or non-preemptive. However, practical scenarios call for a bounded-preemptive scheduling model, which allows to reduce costs and the runtime overhead caused by preemptions. We study the bounded preemptions model in real-time scheduling, where the total number of preemptions, or the number of preemptions per job is bounded by an integer parameter, k. The goal is to feasibly schedule on a single machine a subset of the jobs of maximum total weight. Our theoretical results establish an interesting distinction between the hardness status of non-preemptive, or (unbounded) preemptive real-time scheduling, and k-bounded preemptive scheduling, already for uniform-length jobs with unit weights, for a wide range of values of k. Constant-factor approximation algorithms are obtained for subclasses of instances. Our experimental study shows that while k-bounded preemptive scheduling is hard to solve already on highly restricted instances, simple priority-based heuristics yield almost optimal schedules for realistic inputs and arbitrary values of k.

3.19 Smarter tools for (Citi)bike-sharing

David Shmoys (Cornell University, US)

New York launched the largest bike-sharing system in North America in May 2013. We have worked with Citibike, using analytics and optimization to change how they manage the system. Huge rush-hour usage imbalances the system; we answer both of the questions: where should bikes be at the start of a rush hour and how can we mitigate the imbalances that develop? We will outline the analytics we have employed for the former question, where we developed an approach based continuous-time Markov chains combined with integer programming models to compute daily stocking levels for the bikes, as well as methods employed for optimizing the capacity of the stations. For the latter problem, we present a 3-approximation algorithm used to target limited available rebalancing resources during rush-hour periods. The goal is to ensure that users are never too far from a station that will be rebalanced when looking for a bike or a spot to return one. We formulate this as a variant of the k-center problem and provide a linear programming-based algorithm with a performance guarantee of 3. Finally, we briefly discuss two further planning questions related to their operations, the placement of a limited number of so-called corrals, and scheduling the maintenance of the batteries powering the stations.

3.20 No-wait scheduling for locks

Frits C. R. Spieksma (KU Leuven, BE)

Joint work of Ward Passchyn and Dirk Briskorn.
Main reference W. Passchyn, D. Briskorn, F. C. R. Spieksma, "Mathematical programming models for lock scheduling with an emission objective", European Journal of Operational Research (EOR), 248(3):802–814, 2016.
URL http://dx.doi.org/10.1016/j.ejor.2015.09.012

We investigate a problem inspired by the practical setting of scheduling a lock with multiple chambers. We show how this problem relates to known interval scheduling problems. In particular, for a lock consisting of two chambers we are able to characterize the feasible instances, and use this result to obtain efficient algorithms. We also provide an efficient algorithm for the special case with identical lock chambers. Furthermore, we describe a dynamic programming approach for the more general case with arbitrary chambers, and prove that the problem is strongly NP-complete when the number of chambers is part of the input.

3.21 Lift-and-Round to Improve Scheduling on Unrelated Machines

Ola Svensson (EPFL – Lausanne, CH)

Joint work of Nikhil Bansal and Aravind Srinivasan
Main reference N. Bansal, O. Svensson, A. Srinivasan, "Lift-and-Round to Improve Weighted Completion Time on Unrelated Machines", arXiv:1511.07826 [cs.DS], 2015.
URL http://arxiv.org/abs/1511.07826v2

Weighted completion time and makespan are some of the most relevant and well-studied measures of quality of service in scheduling and resource allocation problems. While these objectives have been studied since the 50's, a systematic study of their approximability was started in the 90's. Since then, impressive progress has led to a complete understanding of these objectives in simple machine models, such as identical machines.

In contrast, it remains a notorious problem to understand the approximability in the more general unrelated machine setting: the classic algorithms developed in the 90's resisted any improvements while having guarantees that are far from the strongest known lower bounds.

In this talk, we overview recent developments with a focus on our recent approximation algorithm that improves upon the barrier of 3/2 for the weighted completion time objective. The improvement is based on a new lift-and-project based SDP relaxation and a general bipartite-rounding procedure that produces an assignment with certain strong negative correlation properties.

This is based on joint work with Nikhil Bansal and Aravind Srinivasan.

3.22 Closing the gap for makespan scheduling via sparsification techniques

Jose Verschae (Pontifical Catholic University of Chile – Santiago, CL)

Joint work of Klaus Jansen and Kim-Manuel Klein
Main reference K. Jansen, K.-M. Klein, J. Verschae, "Closing the Gap for Makespan Scheduling via Sparsification Techniques", in Proc. of the 43rd Int'l Colloquium on Automata, Languages and Programming (ICALP 2016), 2016; to appear.

Makespan scheduling on identical machines is one of the most basic and fundamental packing problems studied in the discrete optimization literature. It asks for an assignment of n jobs to a set of m identical machines that minimizes the makespan. The problem is strongly NP-hard, and thus we cannot expect a $(1+\epsilon)$-approximation algorithm to have polynomial dependency on $1/\epsilon$. Very recently, Chen et al. showed that the dependency has to be exponential if we assume the Exponential Time Hypothesis. A long sequence of algorithms have been developed that try to obtain low dependencies on $1/\epsilon$, but they are all super-exponential on $1/\epsilon$. In this talk I will discuss an improved algorithm that almost matches the lower bound by Chen et al., essentially closing this long line of research. The new ingredient of our approach is an observation that guarantees the existence of a highly symmetric and sparse almost-optimal solution. This structure can then be exploited by integer programming techniques and enumeration. The same idea helps us to obtain improved algorithms for a number of different related problems, including the minimum makespan problem on related machines.

3.23 Approximation algorithms for geometric packing problems

Andreas Wiese (MPI für Informatik – Saarbrücken, DE)

Joint work of Giorgi Nadiradze and Anna Adamaszek

A common setting in geometric packing problems is that we are given a set of two-dimensional items, e.g., rectangles, and a rectangular container and our goal is to pack these items or a subset of them items into the container to optimize objective functions like the total profit of the packed items or the necessary height of the container. A typical obstacle in these problem settings is that in the input there are different types of items, i.e., items that are wide and thin, that are high and narrow, or items that are large in both dimensions. In this talk, I will present a method to handle this obstacle. In a nutshell, the key is to prove that there are near-optimal solutions in which the given container can be partitioned into few rectangular boxes such that in each box there are only items of one of the mentioned types. This leads to better approximation guarantees for two specific problems: a $(1+\epsilon)$-approximation algorithm in quasi-polynomial time for the two-dimensional knapsack problem and a $(1.4+\epsilon)$-approximation algorithm in pseudo-polynomial time for the strip-packing problem. Note that the latter bound is strictly smaller than the lower bound of 3/2 that holds for (non-pseudo-)polynomial time algorithms for the problem.

3.24 Scheduling for Electricity Cost in Smart Grid

Prudence W. H. Wong (University of Liverpool, GB)

Joint work of Mihai Burcea, Wing-Kai Hon, Hsiang-Hsuan Liu, David K. Y. Yau
Main reference M. Burcea, W.-K. Hon, H. H. Liu, P. W. H. Wong, D. K. Y. Yau, "Scheduling for Electricity Cost in Smart Grid", in Proc. of the 7th Int'l Conf. on Combinatorial Optimization and Applications (COCOA 2013), LNCS, Vol. 8287, pp. 306–317, Springer, 2013; pre-print available from author's webpage.
URL http://dx.doi.org/10.1007/978-3-319-03780-6_27
URL https://www.cs.purdue.edu/homes/lans/20061129/publications/burcea2013cocoa.pdf

We study a scheduling problem arising in demand response management in smart grid. Consumers send in power requests with a flexible feasible time interval during which their requests can be served. The grid controller, upon receiving power requests, schedules each request within the specified interval. The electricity cost is measured by a convex function of the load in each timeslot. The objective is to schedule all requests with the minimum total electricity cost. We study different variants and obtain exact algorithms and online algorithms.

4 Open problems

4.1 Multiprocessor Sleep-State Scheduling

Antonios Antoniadis

Consider a set of processors each of which has a unit power consumption per time unit while being active and can transition to sleep at a fixed energy cost α. We are given a set of tasks each associated with a release time, a deadline and a length, and wish to preemptively schedule them so that we minimize one of the following two objectives: (i) The total power consumption. (ii) The number of gaps. Here, a gap is defined as a maximal interval during which a processor is not processing any task. ((ii) can be seen as a special case of (i).)

Questions:
1. Are the two problems NP-hard? Polynomial-time solvable?
2. Develop a constant-factor approximation algorithm for any of the problems?
3. With respect to (ii), is the Longest Viable Gap (LVG) algorithm a constant factor approximation algorithm? Roughly speaking, LVG is a greedy algorithm that repeatedly identifies and fixes the longest possible gap in the instance.

4.2 Exact Speedup Factor for Fixed Priority Pre-emptive versus Fixed Priority Non-pre-emptive Scheduling

Rob I. Davis

The Speedup Factor comparing two real-time scheduling algorithms X and Y is given by the minimum factor by which the speed of the processor needs to be increased so as to ensure that any task set that is schedulable under algorithm Y is guaranteed to also be schedulable under algorithm X. Earliest Deadline First Pre-emptive (EDF-P) scheduling is an optimal single processor scheduling algorithm and so dominates Fixed Priority Pre-emptive (FP-P) and Fixed Priority Non-Pre-emptive (FP-NP) scheduling. There is; however, no such dominance relationships between fixed priority pre-emptive and fixed priority non-pre-emptive scheduling. Determining the exact speedup factor for FP-P v. FP-NP is the focus of this talk. The scope of the problem is uniprocessor systems, assuming the sporadic (real-time) task model. We show, via an example, that the speedup factor is lower bounded by $\sqrt{2}$. We then present the results of experiments using a genetic algorithm to search for high speedup factor task sets. These results indicate that the speedup factor appears to also be upper bounded by $\sqrt{2}$. We conjecture that the exact speedup factor is $\sqrt{2}$. Proof is needed!

4.3 Approximate Mixed-Criticality Scheduling

Christoph Dürr

We consider a single machine scheduling problem, where we want to minimize the makespan, but the exact processing times of jobs is not known at the moment we commit to the starting times of the jobs. But we have the possibility at execution time to drop some less important jobs when the long execution of some previous jobs prevent their executions. This problem falls into the area of mixed criticality scheduling, which has been studied mostly in the real time scheduling community [1, 2].

Formally for this problem there is a single machine and k levels of criticality[1]. Every given job j has a priority level $\ell_j \in [k]$ and processing times $p_j^1 \leq \ldots \leq p_j^{\ell_j}$.

The goal is to produce a schedule defined by starting times S_j for each job, such that for every criticality $c \in [k]$ and jobs i, j with $\ell_i, \ell_j \geq c$ we have $S_i + p_i^c \leq S_j$ or $S_j + p_j^c \leq S_i$. The goal is to minimize the *makespan*, which is defined as $\max_{c\in[k]} \max_{j:\ell_j \geq c} S_j + p_j^c$.

job j	priority level	p_j^1	p_j^2	p_j^3
1	3	2	3	5
2	1	1		
3	2	1	2	

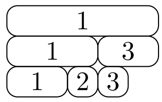

Figure 1 Example of an optimal schedule: the rows correspond to criticality levels and the columns to time slots.

The problem could be strongly NP-hard as for $k = 2$ by a possible reduction from the 3-Partition problem. What is the best achievable approximation ratio by a polynomial time algorithm? Could it be a constant?

In [3] a greedy algorithm is analyzed for the special case $p_j^c = c$, and showed to have ratio between 1.05 and 1.5. It should be possible to tighten the bounds for this special case.

References

1 S. Baruah, V. Bonifaci, G. D'Angelo, H. Li, A. M. Spaccamela, N. Megow, and L. Stougie. Scheduling real-time mixed-criticality jobs. *IEEE Transactions on Computers*, 61(8):1140–1152, 2012.
2 A. Burns and R. I. Davis. Mixed criticality systems – a review. 7th edition, 2016.
3 C. Dürr, Z. Hanzálek, C. Konrad, R. Sitters, O. C. Vásquez, and G. J. Woeginger. The triangle scheduling problem. *CoRR*, abs/1602.04365, 2016.

[1] $[k]$ stands for the set $\{1, \ldots, k\}$.

4.4 Scheduling and Load Balancing with Recourse

Anupam Gupta and Amit Kumar

In many online scheduling settings, we require jobs to be dispatched immediately to a machine. Consider the example of load balancing. Jobs arrive online, and each job j specifies a size p_j and a subset S_j of machines in which it can be scheduled (the so-called "restricted assignment" model). When a job arrives, it needs to be dispatched to a machine. The goal is to minimize the maximum load on a machine, where the load on machine is defined as the total processing time of jobs dispatched to it. The greedy algorithm, which dispatches a job to the least loaded machine, has competitive ratio of $O(\log m)$, where m is the number of machines; and it is known that this is tight even when all jobs sizes are same [2]. If we were allowed to change the allocation of jobs to machines (instead of immediate dispatch), we could solve the off-line load balancing problem at each time instant. However, this may completely disrupt the current assignment of jobs to machines. In the scheduling with recourse model, we ask the following question : Can we improve the competitive ratio if we are allowed to make small changes in the job assignment? As an example, we can show the following: consider the load balancing problem described above in the restricted assignment setting with unit size jobs. We can get a constant competitive algorithm if we are allowed to change the assignment of one existing job (in an amortized manner) whenever a new job arrives [3]. In general, we seek to study how much advantage recourse gives us over traditional competitive analysis. Here are some specific questions:

- Consider the setting of load balancing as above for general job sizes. Can we get constant competitive algorithms with constant amount of job reassignment per job arrival? We can only get an online algorithm which is O(log log n)-competitive (with constant recourse) [3].
- Suppose jobs are arriving over time (in restricted assignment setting), and we want to minimize the maximum flow-time of a job. If we insist on immediate dispatch, it is known that there is no constant competitive algorithm [1]. What happens when we allow recourse?

References

1 S. Anand, K. Bringmann, T. Friedrich, N. Garg, and A. Kumar. Minimizing maximum (weighted) flow-time on related and unrelated machines. In *Proceedings of the 40th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 13–24, 2013.

2 Y. Azar, J. S. Naor, and R. Rom. The competitiveness of on-line assignments. In *Proceedings of the Third Annual ACM-SIAM symposium on Discrete Algorithms (SODA)*, pages 203–210, 1992.

3 A. Gupta, A. Kumar, and C. Stein. Maintaining assignments online: matching, scheduling, and flows. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 468–479, 2014.

4.5 Wireless Scheduling

Magnus M. Halldórsson

Let $p_1, p_2, \dots, p_n$ be points on the real line with *capacities* $c_1, \dots, c_n$. The problem is to partition $P = \{p_i\}$ into fewest sets $P_1, \dots, P_t$, such that

$$\sum_{p' \in P_i, p' \neq p} |p - p'|^3 \leq c_i, \qquad \text{for each } i \text{ and each } p \in P_i.$$

We seek a $O(1)$-approximation. This problem statement captures the most basic open question in scheduling wireless links under the physical (or, SINR) model. Normally, links are given as sender-receiver pairs, but it is known that when messages are all transmitted with the same uniform power, we can blur the distinction between sender and receiver, by paying a constant factor. The problem is usually specified on the plane, or in a general distance metric, but results for the one-dimensional case can typically be generalized relatively easily. The exponent "3", known as the path-loss constant, is situation dependent, and can be any number between 2 and 6.

A $O(1)$-approximation is known for the throughput problem of finding a single set P_1 of maximum cardinality within which all points satisfy the inequality above [1]. This immediately give a $O(\log n)$-factor, but no better is known.

References

1 O. Goussevskaia, R. Wattenhofer, M. M. Halldórsson, and E. Welzl. Capacity of arbitrary wireless networks. In *28th IEEE International Conference on Computer Communications (INFOCOM)*, pages 1872–1880, 2009.

4.6 Stochastic and Multi-Dimensional Scheduling, Generalized Min-Sum Set Cover

Sungjin Im

1. Stochastic scheduling. We are asked to schedule jobs on parallel identical machines non-preemptively with the goal of minimizing total completion time in expectation when the probability distribution on the size of each job is known a priori. The distributions can be arbitrary. Is there a constant approximation for this problem? Even constant approximations with $O(1)$ machine augmentation would be interesting.
2. Generalized min-sum set cover. There is a collection of sets over a universe of elements. Each set S_i has a requirement k_i. We are asked to order elements linearly. A set is said to be covered at time t if t is the first time t' such that k_i elements in S_i appear in the first t' elements of the ordering. The goal is to minimize the total cover time of all sets. The question is if there exists a 4-approximation for this problem. When $k_i = 1$ for all i, there are 4-approximations known. If $k_i = |S_i|$ for all i, there are 2-approximations known. But for the general k_i, the best known approximation factor is greater than 12.
3. In multi-dimensional scheduling, each job j has a demand vector over multiple resources r_j. One can process jobs at rates of $\{x_j\}$ if $\sum_j r_j x_j \leq 1$. The goal is to minimize total

flow time. A scalable algorithm is known for this problem, but the competitive ratio has an exponential dependency on D where D is the number of dimensions. Is it possible to get a $O(1)$-speed $O(poly(D))$-competitive algorithm?

4.7 Generalized assignment

Christos Kalaitzis

The Generalized Assignment Problem is the problem of assigning items, which have a certain size and value, to unit-capacity bins , such that the capacity of the bins is respected, and the sum of the values of the assigned items is maximized. For this problem, it is known that an LP-relaxation called the Configuration-LP has an integrality gap of $1 - 1/e + c$, for some small $c > 0$, which also constitutes the best approximation ratio we know. It would be interesting to look into specific restrictions of the problem, and come up with new techniques to tighten the gap between the upper and lower bounds we have for the Configuration-LP.

The Restricted Max-Min Allocation Problem is the problem of assigning jobs (each of which has a certain processing time) to machines (where each job can only be assigned to a specific subset of the machines), such that all jobs are processed and we maximize the minimum of the sum of processing times over all machines. For this problem, we know that the Configuration-LP has an integrality gap between 1/4 and 1/2, and it would be interesting to investigate its exact value (even if the resulting algorithm is inefficient).

4.8 Flow Time Scheduling on a Line Network

Benjamin Moseley

Say you are given a graph $G = (V, E)$ that is a line. Each node of the graph is a router that forwards packets. Each router can forward up to one packet each time step. Assume that the routers always forward packets to the router to the right for simplicity. There are no buffers and routers can store any number of packets, but can only forward one per time unit.

Requests arrive over time where a request i has an arrival time r_i, a source node s_i and a destination node d_i. Each request is for a single packet to be routed from s_i to d_i. Under an algorithm A, let C_i^A be the time that the packet for request i reaches the designation router d_i. The goal is to decide how packets should be scheduled (routed) to minimize $\sum_i C_i^A - r_i$, the average flow time of the schedule.

See [1, 2] for related work. It is known that no $O(1)$ competitive online algorithm exists. There are no non-trivial upper bounds, even with resource augmentation. A clear goal would be to find a $O(1)$-speed $O(1)$-competitive algorithm. The work of [1] discusses possible algorithms for the problems and their strengths, weaknesses and some properties. The work also gives a $1+\epsilon$ speed $O(1)$-competitive algorithm for the same problem, but the objective is minimizing the maximum flow time. The work of [2] considers the same objective, but looks at tree networks where the source is always the same and gives $O(1)$-speed $O(1)$-competitive algorithms.

References

1 A. Antoniadis, N. Barcelo, D. Cole, K. Fox, B. Moseley, M. Nugent, and K. Pruhs. Packet forwarding algorithms in a line network. In *Proceedings of the 11th Latin American Symposium (LATIN)*, pages 610–621, 2014.

2 S. Im and B. Moseley. Scheduling in bandwidth constrained tree networks. In *Proceedings of the 27th ACM on Symposium on Parallelism in Algorithms and Architectures (SPAA)*, pages 171–180, 2015.

4.9 Are Existentially Scalable Algorithms Really Necessary?

Kirk Pruhs

A scheduling algorithm A is universally scalable if it is $(1+\epsilon)$- speed $O_\epsilon(1)$-competitive, where the subscript of ϵ denotes that the constant can depend upon ϵ. A collection $\{A_\epsilon\}$ of scheduling algorithms, one for each $\epsilon > 0$, is existentially scalable if each A_ϵ is $(1+\epsilon)$-speed $O_\epsilon(1)$-competitive. Thus in this collection the algorithm that 1.1-speed $O(1)$-competitive may well not be the same algorithm that is 1.01-speed $O(1)$-competitive. There are many online scheduling problems where existentially scalable algorithms are known, but no universally scalable online algorithm is known. Examples include broadcast scheduling to minimize total flow time [1], scheduling jobs with arbitrary speed-up curves to minimize total flow[3], broadcast scheduling to minimize maximum flow [2], scheduling unrelated machines to minimize p-norms of flow [4]. A natural, and I believe important, open question is whether universally scalable algorithms exist for these problems. In this short talk, I want to publicize this problem, and briefly discuss some of the difficulties one immediately encounters.

References

1 N. Bansal, R. Krishnaswamy, and V. Nagarajan. Better scalable algorithms for broadcast scheduling. *ACM Transactions on Algorithms (TALG)*, 11(1):3, 2014.

2 C. Chekuri, S. Im, and B. Moseley. Online scheduling to minimize maximum response time and maximum delay factor. *Theory of Computing*, 8(1):165–195, 2012.

3 J. Edmonds and K. Pruhs. Scalably scheduling processes with arbitrary speedup curves. *ACM Transactions on Algorithms (TALG)*, 8(3):28, 2012.

4 S. Im and B. Moseley. An online scalable algorithm for minimizing ℓ_k-norms of weighted flow time on unrelated machines. In *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms(SODA)*, pages 95–108, 2011.

4.10 Precedence-Constrained Scheduling on Identical Machines to minimize weighted completion time

Ola Svensson

To minimize weighted sum of completion times in the presence of precedence constraints is well understood in the case of a single machine: there are plenty of 2-approximation algorithms and the factor 2 is tight assuming a variant of the UGC.

The natural question that follows is what happens if we consider parallel identical machines? In this case, the best algorithm is due to Queyranne and Schulz who gave a 4-approximation algorithm. However the best known integrality gap of the known LP is 2 which is the hardness of the single machine problem.

So the research question is to design a better algorithm for $P|prec|\sum w_j C_j$ or at least give a stronger integrality gap.

4.11 Scheduling with state-dependent machine speed

Tjark Vredeveld

During the 2013 Dagstuhl seminar on Scheduling, Urtzi Ayesta proposed the following open problem. Given are n jobs that need to be processed on a single machine. The jobs have a processing time p_j and weight w_j. The machine may preempt the job. Moreover, the speed of the machine is dependent on the number of jobs in the system. Hereto, speeds $s_1, \ldots, s_n$ are given, where speed s_i denotes the speed of the machine when $i-1$ jobs have been completed. The goal is to minimize the total weighted completion time.

When all speeds are equal, $s_i = 1$ for all $i = 1, \ldots, n$, then the preemption is not used and this scheduling problem is solved by Smith's rule [1], also known as the WSPT rule: process jobs in order of non-increasing ratio $w_j = p_j$. On the other hand, for arbitrary varying speeds, preemption may be useful. The preemption model can be viewed as a processor sharing model in which the processing capacity of the machine is divided over several jobs that can be processed simultaneously.

In [2], we proved that there exists an optimal schedule that processes jobs in groups. That is in between two consecutive (different) completion times, the machines processes on a group of one or more jobs that all finish and start at the same time. We furthermore proved that the general problem is strongly NP-hard by a reduction from 3-Partition. In the same paper, we also designed a greedy algorithm that, given the order of job completions, decides the optimal grouping of jobs. That is, the main issue is to determine the right order in which the jobs should complete.

Even in the case of monotone non-decreasing or non-increasing speeds, WSPT does not necessarily give an optimal order on the job completions. This can be seen by the following two examples.

For the case of non-decreasing speeds, the example consists of 2 jobs. The first job has processing time $p_1 = \epsilon$ and weight $w_1 = 0$. The second job has processing time $p_2 = A$ and weight A, for some large A. The speed of the machine, when two jobs are still in the system, is $s_1 = 1$ and when one job has completed the speed becomes $s_2 = A$. Having the jobs finished in WSPT order, we first process job 2 and then job 1, giving a total weighted completion time of WSPT $= A^2$. The optimal ordering would be to finish first the small and unimportant job 1, so that the large job 2 can be processed at speed A. Then the total weighted completion time will be $\mathsf{OPT} = A(1+\epsilon)$. Hence, $\frac{\mathsf{WSPT}}{\mathsf{OPT}} \geq A - \epsilon' = \frac{s_{\max}}{s_{\min}} - \epsilon'$. By giving the optimum the power to process all jobs at maximum speed and WSPT the disadvantage of processing all jobs at minimum speed, it is easy to see that it also holds that for all instances $\frac{\mathsf{WSPT}}{\mathsf{OPT}} \leq \frac{s_{\max}}{s_{\min}}$.

For the case of non-increasing speeds, the bound is better. A lower bound can be given by the following example of two jobs. The first job has processing time and weight $w_1 = p_1 = 1$

and the second job has processing time $w_2 = 1 + \sqrt{2}$ and $p_2 = w_2 + \epsilon$. Hence, in the WSPT order job 1 precedes job 2. The speeds of the machines are $s_1 = \sqrt{2}$ and $s_2 = 1$. It can be shown that when having the jobs finished in WSPT order, it is equally good to first process job 1 and then job 2 as processing both jobs at the same time. This results in a total weighted completion time of $\mathsf{WSPT} = 4 + 3\sqrt{2}$. The optimum would first process job 2 and then job 1, which gives a total weighted completion time of $\mathsf{OPT} = 4 + 2\sqrt{2}$. Hence, we have a lower bound on the approximation ratio of WSPT of $\frac{\mathsf{WSPT}}{\mathsf{OPT}} \geq \frac{1+\sqrt{2}}{2}$. The algorithms that processes all jobs in one group gives a 2-approximation for the case of non-increasing speeds.

We are interested in the following open problems.

- The NP-hardness proof uses the fact that the processing times may change arbitrarily. What is the complexity of the problem when the speeds are monotone non-decreasing or non-increasing?
- What is the complexity of the non-preemptive version of the problem?
- What is the approximation ratio of WSPT for the case of non-increasing speeds?

References

1. W. E. Smith. Various optimizers for single-stage production. *Naval Research Logistics Quarterly*, 3(1-2):59–66, 1956.
2. V. Timmermans and T. Vredeveld. Scheduling with state-dependent machine speed. In *Workshop on Approximation and Online Algorithms (WAOA)*, pages 196–208, 2015.

Participants

- Fidaa Abed
TU München – Munich, DE
- Susanne Albers
TU München, DE
- Antonios Antoniadis
MPI für Informatik – Saarbrücken, DE
- Yossi Azar
Tel Aviv University, IL
- Nikhil Bansal
TU Eindhoven, NL
- Sanjoy K. Baruah
University of North Carolina at Chapel Hill, US
- Vincenzo Bonifaci
CNR – Roma, IT
- Niv Buchbinder
Tel Aviv University, IL
- Marek Chrobak
University of California – Riverside, US
- Bouke Cloostermans
TU Eindhoven, NL
- Liliana Cucu-Grosjean
INRIA – Paris, FR
- Robert Davis
University of York, GB
- Christoph Dürr
UPMC – Paris, FR
- Thomas Erlebach
University of Leicester, GB
- Anupam Gupta
Carnegie Mellon University, US
- Magnus M. Halldórsson
Reykjavik University, IS
- Sungjin Im
University of California – Merced, US
- Klaus Jansen
Universität Kiel, DE
- Christos Kalaitzis
EPFL Lausanne, CH
- Samir Khuller
University of Maryland – College Park, US
- Amit Kumar
Indian Inst. of Technology – New Dehli, IN
- Retsef Levi
MIT – Camridge, US
- Alberto Marchetti-Spaccamela
Sapienza University of Rome, IT
- Monaldo Mastrolilli
IDSIA – Manno, CH
- Nicole Megow
TU München, DE
- Rolf H. Möhring
TU Berlin, DE
- Benjamin J. Moseley
Washington University – St. Louis, US
- Seffi Naor
Technion – Haifa, IL
- Kirk Pruhs
University of Pittsburgh, US
- Thomas Rothvoss
University of Washington – Seattle, US
- Jiri Sgall
Charles University – Prague, CZ
- Hadas Shachnai
Technion – Haifa, IL
- David Shmoys
Cornell University, US
- René Sitters
VU University of Amsterdam, NL
- Frits C. R. Spieksma
KU Leuven, BE
- Clifford Stein
Columbia University, US
- Ola Svensson
EPFL – Lausanne, CH
- Marc Uetz
University of Twente, NL
- Suzanne van der Ster
TU München, DE
- Rob van Stee
University of Leicester, GB
- Jose Verschae
Pontifical Catholic University of Chile – Santiago, CL
- Tjark Vredeveld
Maastricht University, NL
- Andreas Wiese
MPI für Informatik – Saarbrücken, DE
- Gerhard J. Woeginger
TU Eindhoven, NL
- Prudence W. H. Wong
University of Liverpool, GB

Report from Dagstuhl Seminar 16091

Computational Challenges in Cooperative Intelligent Urban Transport

Edited by

Caitlin Doyle Cottrill[1], Jan Fabian Ehmke[2], Franziska Klügl[3], and Sabine Timpf[4]

1 **University of Aberdeen, GB,** `c.cottrill@abdn.ac.uk`
2 **FU Berlin, DE,** `janfabian.ehmke@fu-berlin.de`
3 **University of Örebro, SE,** `franziska.klugl@oru.se`
4 **Universität Augsburg, DE,** `sabine.timpf@geo.uni-augsburg.de`

Abstract

This report documents the talks and group work of Dagstuhl Seminar 16091 "Computational Challenges in Cooperative Intelligent Urban Transport". This interdisciplinary seminar brought researchers together from many fields including computer science, transportation, operations research, mathematics, machine learning and artificial intelligence. The seminar included two formats of talks: several minute research statements and longer overview talks. The talks given are documented here with abstracts. Furthermore, this seminar consisted of significant amounts of group work that is also documented with short abstracts detailing group discussions and planned outcomes.

Seminar February 28–4, 2016 – http://www.dagstuhl.de/16091
1998 ACM Subject Classification H.4.2 Types of Systems
Keywords and phrases collaborative transport, Computational transportation science, Cooperative computing, Crowd-sourcing of transportation data, Intelligent transportation systems
Digital Object Identifier 10.4230/DagRep.6.2.119
Edited in cooperation with Kevin Tierney

1 Executive Summary

Caitlin Doyle Cottrill
Jan Fabian Ehmke
Franziska Klügl
Sabine Timpf

Following the history of two Dagstuhl seminars on Computational Issues in Transportation in 2010 and 2013, the organizers of this follow-up seminar concentrated on upcoming, data-driven challenges in the area of urban transport. In recent years, urban transportation networks have become more diverse, with a growing mix of public and private operators providing disaggregated services and information. The resulting multitude of transportation options includes non-traditional modes and services such as car and bike sharing in addition to established public transport and individual car options. So far, it is challenging to combine detailed operational data automatically arising from these services, since these data are generated both from service operation and from the users of services via crowdsourcing. The seminar aimed to discuss how data sources can be made available for individual planning and system-wide coordination of urban transportation using an approach from distributed computing, i.e., getting all involved parties to cooperate in providing relevant spatial and temporal information in a timely fashion. It was not clear how to derive reliable information for planning and control approaches, or how to adapt optimization methodologies to make urban transportation more cooperative and intelligent.

The aims of the seminar were to extend the existing network in disciplines such as Computational Traffic Science, Optimization, Autonomic Computing and Artificial Intelligence for discussing computational challenges in cooperative intelligent urban transportation, mesh communities by collecting suggestions for (partial) solutions for burning issues in urban transportation and discussing the prerequisites for merging into interdisciplinary approaches, document the state of the art and current computational challenges in cooperative intelligent transportation.

To this end, an interdisciplinary group from areas such as computer science, geography, applied optimization and traffic engineering met at Dagstuhl. The number of attendees was advantageous for group discussions, not too small for breakout groups but also not too large for meaningful discussions in the plenum.

We started on Sunday evening with a game ("Cards Against Urbanity – special issue for this seminar") specifically designed for this event by Ms. Cottrill. The game was a great success as icebreaker and helped bringing together the participants with their various backgrounds. Monday was opened with a keynote by Vonu Thakuriah, who discussed examples, prospects and challenges of emerging forms of data in transportation research and applications. The participants introduced themselves, bringing a significant object describing their relationship with the seminar's topic.

For the remaining seminar time, the participants were asked to contribute to the seminar's content by one of the following options: they could give an overview talk of an emerging area (20 minutes), a research statement on what they have been working on in their particular area (5 minutes), and they were asked to come together in groups that were defined dynamically on Monday afternoon. The resulting abstracts can be found in this report. Based on the participants' interests, groups discussing the topics of online simulation, pedestrian behavior, autonomous transportation, smart cities, and benchmark data emerged. On Wednesday afternoon, the participants went on a 'field trip' to the retail lab by DFKI in St. Wendel, where the future of retail can be explored hands-on. Since there was a significant interest in the provision of benchmark data for urban transport, there was a special session and group work devoted to this topic on Thursday afternoon. Friday morning was meant for collecting the results of the group work and collecting open challenges for future seminars.

Summarizing, the seminar identified computational challenges to cooperative intelligent urban transport, among others notably research on opportunistic groups in public transport (i.e., people sharing tickets and or trajectories in an ad-hoc fashion), freight pods attached to light rail (i.e., mixing of freight and passenger transportation), define a common language for sharing complex knowledge and real-time data in smart cities and creating benchmark datasets for different modelling purposes and at different scales. We think that the seminar was quite successful in extending the existing networks by bringing together researchers from many different disciplines relevant for the future of urban transport. Some of the groups are planning to write proposals for the appropriate EU calls coming out in October, while others have started to work on position papers describing the state of the art as well as resulting future challenges of the field.

2 Table of Contents

3 Overview of Talks

3.1 Ridesharing and crowdshipping

Niels Agatz (Erasmus University – Rotterdam, NL)

In this presentation I discuss various initiatives that allow people to share rides or ship freight by using journeys that already take place. These systems may provide significant societal and environmental benefits by reducing the number of cars on the road and improving the utilization of available seat capacity. I present the results from several research projects in this area that aim to provide decision support models to match drivers and riders/parcels in different settings.

3.2 Human-centered intelligent transport systems

Ana Lucia Bazzan (Federal University of Rio Grande do Sul, BR)

In this slides I:

- present some motivation for human-centered Int. Transp. Systems
- discuss common issues between multiagent systems, complex systems, and transportation systems
- present an overview of my current projects

3.3 E-Fulfillment for Attended Last-Mile Delivery Services in Metropolitan Areas

Catherine Cleophas (RWTH Aachen, DE)

Our project aims to optimize the final part of a firm's value chain with regard to attended last-mile deliveries. We assume that to be profitable, e-commerce businesses need to maximize the overall value of fulfilled orders (rather than their number), while also limiting costs of delivery. To do so, it is essential to decide which delivery requests to accept and which time windows to offer to which consumers. This is especially relevant for attended deliveries, as delivery fees usually cannot fully compensate costs of delivery given tight delivery time windows. Existing order acceptance techniques often ignore either the order value or the expected costs of delivery.

We present an iterative solution approach: after calculating an approximate transport capacity based on forecasted expected delivery requests and a cost-minimizing routing, actual delivery requests are accepted or rejected aiming to maximize the overall value of orders given the computed transport capacity. With the final set of accepted requests, the routing solution is updated to minimize costs of delivery. The presented solution approach combines well-known methods from revenue management and time-dependent vehicle routing.

References

1 C. Cleophas and J. F. Ehmke *When Are Deliveries Profitable? – Considering Order Value and Transport Capacity in Demand Fulfillment for Last-Mile Deliveries in Metropolitan Areas*. Business & Information Systems Engineering: Vol. 6: Iss. 3, pp. 153–163, 2014.

3.4 Linking urban infrastructure systems: Future smart cities

Sybil Derrible (University of Illinois – Chicago, US)

Transportation systems are but one of the many infrastructure systems that populate cities. Whether it is the water/wastewater systems, the electric grid, the natural gas system, the telecom lines, or even the buildings, all urban infrastructure systems are intrinsically interconnected and interdependent. Computational challenges that the transportation sector have to face are in fact shared across most infrastructure systems. In this research statement, I discuss how urban infrastructure systems are physically and operationally linked and how we can better integrate them in the future, notably by presenting the concept of an integration-decentralization matrix and by giving specific examples. Overall, this research argues for a better coordination and planning of urban infrastructure systems, contributing to the future smart city.

3.5 The Computer Systems Group at Universidade Nova de Lisboa

Cecilia Gomes (New University of Lisbon, PT)

The distributed nature of Computational Transportation Science (CTS) and its requirements for large-scale data management and intensive computing raise a set of problems matching the area of Computer Systems. Cooperative Intelligent Transport Systems (CITS), in particular, benefit from novel solutions in mobile technologies and social networks and heavily depend on advances pertaining the big data problem. Namely, data management in all of its dimensions is very present in CITS including data acquisition, data storage and access, data processing and mining, or data dissemination and sharing.

Data acquisition is increasingly made from multiple heterogeneous data sources like vehicle and infrastructure sensors, satellite sources, user's mobile devices, services supported by transport companies and authority entities, etc.

The generated data may range from raw/unprocessed data in real time to off-line data generated by simulations building an historic view of a particular transport system. Such diverse data at different levels of the Intelligent Transport System (ITS) infrastructure may need to be accessed at different times or together requiring flexible forms of data access and uniformization in order to allow related information extraction. For instance, deployed networks of wireless sensors for weather conditions with restricted autonomy may be accessed only for severe weather conditions to be combined with traffic flow data (including from users) in a increasingly congested area. Moreover, CITS require support for data integrity, reliability and security, to guarantee that the acquired data is in fact dependable and not permeable to malicious information injection or extraction.

Such huge amount of produced data requires hence efficient, affordable, and secure ways of data storing, access, and processing. This includes traditional ones like proprietary data centers, to more recent private and hybrid Clouds or public Clouds' upscale solutions in a pay as you need model. Whereas local authorities and established urban transport networks traditionally rely on private centralized systems for data processing and mining, novel transport applications in logistics or car sharing/hiring benefit from novel forms of dependable distributed processing supported by a combination of mobile devices with cost effective Cloud solutions. In the former case, data processing requires traditional high performance computing solutions but also, for instance, easy forms of applications' portability to novel parallel architectures. Computer nodes composed of CPUs and GPU nodes are increasingly present in diverse computational platforms offering economies of scale but still demanding substantial effort on application's deployment.

In the latter case, data processing may be done on transit exploring mobile devices' capabilities but requiring efficient mechanisms on distributed data consistency and security and on using Cloud services whenever necessary (e.g. to locate a novel transportation application's service and its state).

As for data dissemination and sharing, CITS depend on timely data distribution to support decentralized information and service provision. Additionally, they will also increasingly need to explore forms of data sharing and dissemination (e.g. building common contexts like in social networks) among different urban transportation networks and users. For instance, internet and mobile companies may explore the information on travelers' location in order to use efficient broadcast for data dissemination for a particular area or application instead of more inefficient point to point/client-server interaction models. Likewise, different transportation networks may agree on sharing information (e.g. public transports and logistics operators with pre-defined schedules) whenever that may be economical viable.

Towards the above requirements in CITS, the Computer Systems Group at the New University of Lisbon, Portugal, is researching a set of topics that may contribute to this area. The mission of the group is to make the computer systems that surrounds us in our daily lives more reliable, trustworthy, dependable, and better performing following a principled engineering approach. This encompasses designing new algorithms for building systems that address practical needs and using formal methods to reason about that design, and implementing software prototypes for experimentally evaluating those systems. Some of the research areas are Highly-scalable Adaptable Data Dissemination, Decentralized Data Processing, Secure Data Management, Structured Parallel Programming.

In popular applications it is expected that a huge number of users a) Contribute with sensor data, e.g. via their mobile phones, and b) request information from the system. In this scenario, scalability of data processing (streams) becomes a major issue. Namely, a single server is insufficient to process all data and reply to all queries, but data processing may be partitioned among the nodes. The proposed solution, named C4S (Cloud for Sensing) provides a system for supporting participatory sensing applications that leverages cluster or cloud infrastructures to provide a scalable data processing infrastructure. Several decentralized processing strategies for data processing based on geographic partitioning are offered providing different tradeoffs on reliability and processing latency.

Following an increasing trend on moving processing the network edge, the group is also exploring other solutions combining mobile computing and edge Clouds for collaborative computations on mobile devices. The research encompasses forms of mobile to mobile device collaboration where no access to the Cloud is available with computations offloaded to the Cloud when possible. The solution minimizes network usage and significantly reduces latency, building systems that are fast when possible and consistent when needed.

Abstract-Gossip, or epidemic, protocols have emerged as a highly scalable and resilient approach to implement several application level services such as reliable multicast, data aggregation, publish-subscribe, among others. All these protocols organize nodes in an unstructured random overlay network. The goal of this work is to support the decentralized management of these networks supporting high churn tolerance and high scalability, while guaranteeing an efficient, robust and collaborative data dissemination among the nodes. Applications relying either on mobile networks or vehicular networks (e.g. autonomic vehicles in a city or groups of vehicle/user combination of computational devices) may benefit from these type of protocols for efficient and flexible data dissemination.

The group also addresses reliability and privacy-preserving of data via secure data management solutions. Low overhead solutions are being developed that offer secure data search on the cloud since they operate over encrypted data. They provide decentralized data sharing over an high number of replicas and rely on pairwise synchronization, and may be applied to sensitive information like medical data or to perform search over multimedia data repositories. Such secure solution may hence be a good contribution to transportation data processing whenever privacy concerns have to be addressed.

Finally, concerning the need to simplify the programming of the novel computational infrastructures composed of heterogeneous components (multi-core and many-core) the group is contributing to the area of structured (i.e. pattern-based) parallel programming. This area tackles a software engineering approach for (parallel/distributed) applications' development that aims to reuse expert knowledge on programming distributed and parallel applications via pattern abstractions. The goal is to increase the productivity of non-experts in those areas by promoting application portability and reusability in a simpler way.

Relevant publications:

- Scalable Data Processing for Community Sensing Applications. S. Duarte, et al. MONET 2013.
- Pixida: Optimizing Data Parallel Jobs in Wide-Area Data Analytics. K. Kloudas, et al. VLDB'16.
- HyParView: a membership protocol for reliable gossip-based broadcast. J. Leitão, J. Pereira and L. Rodrigues, DSN'07.
- X-BOT: A Protocol for Resilient Optimization of Unstructured Overlay Networks. J.
- Leitão, J. P. Marques, J. Pereira, and L. Rodrigues, IEEE TPDS 2012. Privacy-Preserving Content-Based Image Retrieval in the Cloud. B. Ferreira, J. Rodrigues, J. Leitão, and H. Domingos. SRDS'15.
- Extending Grid-Based Workflow Tools With Patterns/Operators. M.C. Gomes, O.F. Rana, J.C. Cunha. IJHPCA 22(3) (2008)
- Evaluating Novel Parallel Programming Models in the Context of Digital Breast Tomosynthesis. P. Somsen, P. Ferreira, N. Matela, P. Medeiros, and M.C. Gomes. (under submission)

3.6 Population-level filtering of abundant data

Benjamin Heydecker (University College London, GB)

We work in an era when data of one kind or another are abundant. However, the sources and quality of the data are often different from what we would have chosen. Issues that arise include implicitly selective sampling, incomplete observations and erroneous measurements. If we are to use data with these characteristics, then recognising them in our analysis and usage is potentially important.

This presentation will show how analysis of abundant datasets can lead to probabilistic identification and treatment of spurious data. This establishes an approach to filtering data that takes advantage of the volume of data available. The result of this is a probabilistic classification of individual observations as relevant or not, and a consequent approach to filtering information at the population rather than the individual level. An example application to travel time data derived automatically from matched number plate readings shows how this approach can be used to estimate a distribution of travel times under normal conditions alongside a distinct distribution of together with those under adverse conditions. A separate component distribution is estimated for spurious observations, which is taken into account when estimating travel times and when classifying individual observations.

The approach presented is applicable to data streams that are so abundant that detailed analysis of individual observations is impractical.

3.7 Enhancing environmental awareness using Participatory Sensing and the Social Web

Andreas Hotho (Universität Würzburg, DE)

The Social Web allows millions of users to share and exchange information about their daily activities. Smartphones are one way to do this. Even more they become a device to sense the environment and to inform users about the current situation. Thus, subjective data like human perceptions and impressions expressed in the Social Web can easily be combined with additional low cost sensor information, which is an important next step on the way towards establishing the ubiquitous web. The active contribution of everyone carrying such devices as a sensor is known as participatory sensing and allows for the collection of huge masses of new kinds of sensing data, especially as the data most often contains location information. Combining those with Social Web information adds additional value to it and researchers, especially from machine learning, data mining, and social network analysis, are interested in such data. They work on novel methods to understand the underlying human relationship, perception and interactions with the environment.

In the Everyaware project, we made a first steps towards turning a smartphone into a sensor for noise and air pollution measurements. Particularly we combined it with effective low cost air quality sensors which enables everyone to measure the black carbon concentration. More specifically, the developed box allows to determine the personal BC exposure in an urban environment by measuring indicator gases. Additionally users are able to react to the presented measurements using typical social web channels but also simple keyword or hashtags. As time and spatial information are connected with each measurement, the data can be used as a basis for a collective view on the urban environment. Results can be shown in an aggregated form on maps which allows for better insights into the current noise and air pollution situation. In several case studies we collected a lot of measurements alongside with subjective impressions. This data demonstrates the power of the approach and give insights into temporal and spatial distribution of both the noise and the black carbon concentration. The added personal feedback of the users show their perception of the measured environment and their learning behaviour about the perceived perception of their surrounding.

Details about the project results can be found in:

- Sîrbu, A., Becker, M., Caminiti, S., De Baets, B., Elen, B., Francis, L., Gravino, P., Hotho, A., Ingarra, S., Loreto, V., Molino, A., Mueller, J., Peters, J., Ricchiuti, F., Saracino, F., Servedio, V. D. P., Stumme, G., Theunis, J., Tria, F. & Van den Bossche, J. (2015). Participatory Patterns in an International Air Quality Monitoring Initiative. PLoS ONE, 10, e0136763. DOI: 10.1371/journal.pone.0136763
- Becker, M., Caminiti, S., Fiorella, D., Francis, L., Gravino, P., Haklay, M. (M., Hotho, A., Loreto, V., Mueller, J., Ricchiuti, F., Servedio, V. D. P., Sîrbu, A. & Tria, F. (2013). Awareness and Learning in Participatory Noise Sensing. PLoS ONE, 8, e81638. DOi: 10.1371/journal.pone.0081638
- http://cs.everyaware.eu/event/overview
- http://www.everyaware.eu/

3.8 Earliest Arrival Problem of Pedestrians

Tobias Kretz (PTV AG – Karlsruhe, DE)

Problem Formulation

Pedestrians can choose their walking direction continuously. A single pedestrian – undisturbed by other pedestrians – who wants to arrive as early as possible at some given destination would simply walk the shortest path – a poly line that leads around obstacles to destination. If there are other pedestrians around evading these might be required and the actually walked trajectory will deviate from the shortest path. If there is a considerable number of pedestrians around one may adopt a macroscopic perspective and by some given fundamental diagram assign spots continuously an expected walking speed according to the local density. The task is to simulate the behavior of a pedestrian who wants to move in minimum time through such an environment with two continuous spatial dimensions and walking speeds varying over time and space; or – a slightly different task: to find at each moment from continuous 360° the walking direction which momentarily promises the earliest arrival at a given destination.

Solution Idea

To compute estimated remaining travel time to destination one can place a fine (finer than body dimensions) grid over the geometry and assign each grid cell a small estimated time to pass over it. If – beginning from the destination – all these small times are summed up one ends up with a map of estimated remaining travel time to destination. The negative gradients of such a field (resp. map) point into the direction of estimated earliest arrival at destination. Since the normalized gradients are the desired result of the computation and since the normalized gradients are invariant to a global factor of the map of remaining travel times one may arbitrarily define that the estimated time to move over an empty cell of solid ground and even surface be 1. An empty cell with a ground that hinders movement (for example sand) might be 1.5 throughout the simulation. A cell that is occupied by another pedestrian can be chosen to have the value $1 + g$ with g being a parameter that needs calibration. Obviously not all pedestrians pose an equal hindrance for future movement. If some other pedestrian is ahead of "me", but is walking faster than "I" would ideally like to move, then this other pedestrian is not a hindrance. There is no need to begin overtaking. Therefore the "$1 + g$" is extended to the following formula:

$$\frac{1}{f} = 1 + \max\left(0, g\left(1 + h\frac{\vec{v} \cdot \nabla S}{v_0|\nabla S|}\right)\right) .$$

Here h is another parameter for calibration whose value is expected to lie in [0..1]. S is the field of (static) spatial distance to destination. v is the velocity of the pedestrian occupying the spot and v_0 is the desired speed of the pedestrian for whom the field of remaining travel time is to be calculated. Thus, v and v_0 are NOT from the same pedestrian. If the field is calculated to be used by more than one pedestrian v_0 can be chosen to be the average of desired walking speeds.

The resulting $1/f$ is used as right side of the Eikonal Equation

$$|\nabla T(\vec{x})|^2 = \frac{1}{f(\vec{x})^s} .$$

Which then is solved numerically using for example the Fast Marching Method (Kimmel & Sethian, 1998) or the Fast Iterative Method (Jeong & Whitaker, 2007). This mathematical integration is also an integration in the sense that it integrates the effect of all pedestrians as well as the attractive effect of the destination such that the resulting walking direction is no longer a greedy optimization with regard to distance reduction to destination. This implies that adding or taking one single pedestrian to or from a scene may change the effect of all other pedestrians on the one pedestrian under consideration. In social systems "...the properties of any part are determined by its membership in the total functional system." (Sherif, 1936). The desired walking directions can then be computed as

The method as described to this point has been implemented in 2011 and is available in the microscopic simulation PTV Viswalk to make pedestrians walk into the direction of estimated earliest arrival. For this the resulting desired direction is used as direction of the desired velocity in an implementation of the Social Force Model (Johansson, Helbing, & Shukla, 2007).

Computational Challenges

Since the spatial distribution of pedestrians as well as their velocities change within a simulation, the field T has to be computed repeatedly, ideally in each simulation time step. Even if only one field T is computed for all pedestrians heading to a particular destination this is already computationally challenging for typical project size area extent (computation time in this case does not scale with the number of pedestrians, but the area extent covered by T) and current typically available desktop computing power.

As stated ideally each pedestrian would have its own field T which multiplies the effort by the number of pedestrians. As of 2016 this is computationally infeasible to be carried out on typical commercial project size scale.

Another aspect in principle requires a multiple computation time. The gradient of the field S in the first equation has the role of the desired walking direction. However, the desired walking direction is what is intended to be calculated with this method. Thus gradient S is just an approximation to the desired walking direction as input for the method and so T is only an approximation to the field of estimated remaining travel times as output of the method. One could (and ideally would) use the resulting T as input for the same method, replacing S to compute a T_2 and iteratively a T_3, T_4, and so on. Provided the process converges this would be continued until $T_{n+1} - T_n$ is smaller than some given value for all cells.

Further Reading

The solution idea is rolled out in detail in (Kretz, et al., 2011). In (Kretz, 2014) the aspect is discussed that the field T considers the destination and all pedestrians holistically, i.e. that adding one pedestrian may change the effect all other pedestrians have to the pedestrian in focus. The method is used in (Kretz, 2013) to verify the computer implementation of the Social Force Model and itself. In (Kretz, et al., 2011) the parameters of the model are calibrated by means of an immersive virtual reality experiment. Further applications of the method are presented in (Kretz, 2009b) and (Kretz & Große, 2012). Finally in (Kretz, Lehmann, & Hofsäss, 2014) the method is compared to a method which statically computes routing alternatives and distributes pedestrians with an iterative assignment on the routes such that there is an equilibrium of travel times along the routes. A video that animates the mathematical process and shows examples of application can be seen here: https://youtu.be/8SmRBTJ-jeU.

References

1 Jeong, W.-K., & Whitaker, R. (2007). A fast eikonal equation solver for parallel systems. SIAM conference on Computational Science and Engineering, (S. eprint).

2 Johansson, A., Helbing, D., & Shukla, P. (2007). Specification of the social force pedestrian model by evolutionary adjustment to video tracking data. Singapore: World Scientific Publishing.

3 Kimmel, R., & Sethian, J. (1998). Computing geodesic paths on manifolds. PNAS, 95(15), S. 8431–8435.

4 Kretz, T. (2009b). The use of dynamic distance potential fields for pedestrian flow around corners. In T. Delft (Hrsg.), First International Conference on Evacuation Modeling and Management (ICEM). Scheveningen: arxiv:0804.4336.

5 Kretz, T. (2013). Multi-Directional Flow as Touch-Stone to Assess Models of Pedestrian Dynamics. Annual Meeting of the Transportation Research Board 2013, (S. 13–1160).

6 Kretz, T. (2014). The Effect of Integrating Travel Time. In U. Kirsch, U. Weidmann, & M. Schreckenberg (Hrsg.), Pedestrian and Evacuation Dynamics 2012. Zürich: Springer Berlin / Heidelberg.

7 Kretz, T., & Große, A. (2012). From Unbalanced Initial Occupant Distribution to Balanced Exit Usage in a Simulation Model of Pedestrian Dynamics. Human Behaviour in Fire Symposium, (S. 536–540).

8 Kretz, T., Große, A., Hengst, S., Kautzsch, L., Pohlmann, A., & Vortisch, P. (2011). Quickest Paths in Simulations of Pedestrians. Advances in Complex Systems, 14(5), S. 733-759.

9 Kretz, T., Hengst, S., Roca, V., Pérez Arias, A., Friedberger, S., & Hanebeck, U. (2011). Calibrating Dynamic Pedestrian Route Choice with an Extended Range Telepresence System. 2011 IEEE International Conference on Computer Vision Workshops, (S. 166–172).
10 Kretz, T., Lehmann, K., & Hofsäss, I. (2014). User Equilibrium Route Assignment for Microscopic Pedestrian Simulation. In review.
11 Sherif, M. (1936). The Psychology of Social Norms.

3.9 Logistics for Shared Mobility

Dirk Christian Mattfeld (TU Braunschweig, DE)

People's mobility budget seems to be bounded to approximately one hour per day. Urban transportation infrastructure determines the distance that can be covered by travelers within this hour. The greater the flexibility of travel modes, the greater the benefit to inhabitants of a conurbation. Only recently, shared mobility systems offer the individual use of vehicles without required ownership of these vehicles. This flexible use allows for inter-modality and multi-modality of people and directly impacts the reach of daily travel.

However, the availability of shared vehicles will be limited to typical daytime dependent flows of travel within a city. For instance, the demand of commuting from a residential area to a business district in the morning will be high resulting in a competition with respect to vehicles of a shared mobility system. This effect can be shown by the fill level of stations of the bike sharing system of London. Cluster analysis with respect to the bike usage in Vienna conducted by the author reveal similar results.

In order to provide a high level of service in terms of the possible pickup of bikes and the possible return of bikes at stations, a relocation of bikes through the transport operator is necessary. Exemplary figures show, that the Barcelona bike sharing system Bicing requires approximately 4 million Euros for the relocation of bikes per year. This amounts to approximately 1/3 of the operations cost of this system providing access to 6000 bikes to the population of Barcelona, Spain. In order to provide cost efficient operations, a proper planning and control of the system is required by means of logistics optimization models. On the strategic level, the system design of a bike sharing system may be supported by covering models. Transportation models may be used to determine relocation flows of bikes between stations pursuing a balancing of bike numbers among stations. In order to support relocation operations, vehicle routing models may be applied.

Additionally, user incentives with respect to the bike usage may be set by dynamic pricing. Revenue management may contribute to this task in future shared mobility systems. Also, maintenance and repair of shared mobility vehicles may be supported by interval or condition based maintenance in the future. All of these optimization models have been formulated in logistics management and need to be tailored to special needs of shared mobility systems. The actual needs strongly depend on the features of the shared mobility system under consideration.

3.10 Automated Planning

Thomas Leo McCluskey (University of Huddersfield, GB)

Automated Planning, from an artificial intelligence point of view, is a research area in which algorithms are developed that output plans which solve goals. Specifically, planning algorithms input knowledge of actions and resources, and knowledge of an application environment, and given a planning problem, synthesis an ordered set of actions which when executed from a current state will solve the given planning problem.

Planning algorithms developed by the research community have been applied in a wide range of applications, including AUV control, narrative generation and workflow management. In the talk we consider some important areas of research in Automated Planning, and its application to road transport management, where it has proven useful in generating traffic light strategies for urban regions.

3.11 Research Statement: Computer Vision

Andrea Prati (University of Parma, IT)

Within the field of Cooperative Urban Transportation, computer vision (CV) should be considered as an enabling technology. Among the different uses which might be helpful for Urban Transportation, CV can be use as a tool for automatic extraction of (real-time) information about objects (people or vehicles) in an urban scenario. Object detection and tracking from different (live) video streams can be achieved using standard CV algorithms. Object re-identification (i.e., the ability to automatically understand whether an object reappears in a new camera view) might also be employed to keep tracks of object movements (and related statistics, such as travel time, queue formation, incident detection, etc.) on larger areas.

Apart from this extraction of real data (to be compared with or added to simulated data), CV enables also new types of applications related to transportation. Among the many, the implementation of CV algorithms on mobile phones (smartphones and tablets), often called "mobile vision", opens the way to several interesting applications. For example, the automatic place recognition (i.e. the ability to recognize from the camera embedded on the mobile device landmarks in the surroundings) provides a tool for refining the localization of users and vehicles while traveling in a city. An even more innovative application is related to the automatic profiling of users through the use of the frontal-facing embedded camera: mood/emotion, as well as age, gender and ethnicity, could be automatically determined (with a certain accuracy) in order to provide to advanced recommender systems, useful profile information about the user.

Last but not the least relevant, CV algorithms are at the core of autonomous guidance of vehicles, in addition to other sensors.

3.12 Good City Life

Daniele Quercia (NOKIA Bell Labs – Cambridge, GB)

The corporate smart-city rhetoric is about efficiency, predictability, and security. "You'll get to work on time; no queue when you go shopping, and you are safe because of CCTV cameras around you". Well, all these things make a city acceptable, but they don't make a city great. We are launching goodcitylife.org – a global group of like-minded people who are passionate about building technologies whose focus is not necessarily to create a smart city but to give a good life to city dwellers. The future of the city is, first and foremost, about people, and those people are increasingly networked. We will see how a creative use of network-generated data can tackle hitherto unanswered research questions. Can we rethink existing mapping tools [happy-maps]? Is it possible to capture smellscapes of entire cities and celebrate good odors [smelly-maps]?

happy-maps: http://www.ted.com/talks/daniele_quercia_happy_maps
smelly-maps: http://www.di.unito.it/~schifane/smellymaps/project.html

3.13 Overview of Issues/Problems leading to Computational Challenges in Cooperative Intelligent Urban Transport

Jörg-Rüdiger Sack (Carleton University – Ottawa, CA)

In this talk, we set the stage for the seminar by reviewing (see also [1]) some of the main urban transport challenges and identify the resulting tasks:

1. Parking optimization
2. Delivery parking optimization
3. Planning environmentally friendly transport
4. Intelligent city planning
5. Intelligent, cooperative planning of transport mixes
6. Intelligent incorporation of public space
7. Development of software/hardware infrastructure to store, assess, model, plan intelligent infrastructure maintenance, repair, renewal, expansion ...
8. Reduction of accidents on road networks and increase the safety of all participants using cooperative approaches.
9. More and more intelligent city logistics strategies are required taking many environmental factors into consideration: noise, vibration, visual intrusion, ..
10. Develop privacy guaranteeing strategies or clear guideline on extend to which privacy needs to be given up to gain access to services...

We also discuss some of the European Comission work on ITS (see [2]) including the European Union mandates M/45 on C-ITS abd M/453. Finally, we describe some of the obstacles in acceptance of work on C-ITS and an approach how to overcome these.

A proposal we made to TomTom based on work carried out at an earlier Dagstuhl seminar is described. This work is based on an exact algorithm for shortest path in time-dependent

FIFO networks developed by the author and collaborators (former Ph.D. student M. Omran and F. Dehne, Carleton University) and practical approximation algorithms (see [4]).

Task 8 on accidents and safety might benefit from [5], where we describe how to find k "shortest paths" between s and d with the minimum number of shared edges. Distributing traffic between two nodes of the road network is this fashion could result in a better road utilization, reduce traffic and thus accident while having a small additional cost when compared to the cost of an optimal route.

Finally, the important topic of privacy is discussed which is of importance throughout the remainder of the seminar. The question is raised if cooperation is a means to, or an automatic loss of, privacy. Work by the author [**?**] generalizing k-privacy to (i, j)-privacy might be usable to determine if certain information to be shared with others needs to be delayed prior to its sending. We wish to guarantee that even multiple sharing of information along a route taken does not lead to a loss of location privacy.

References

1 http://people.hofstra.edu/geotrans/eng/ch6en/conc6en/ch6c4en.html
2 http://ec.europa.eu/transport/themes/its/index_en.html
3 F. Dehne, M. T. Omran, and J.-R. Sack, "Shortest Paths in Time-Dependent FIFO Networks", Algorithmica (2012) 62:416-435; DOI: 10.1007/s00453-010-9461-6.
4 M. Omran and J.-R. Sack, "Improved Approximation for Time-Dependent Shortest Paths", in Proc. 20th Int. Computing and Combinatorics Conf., COCOON, Atlanta, USA, Aug. 2014, LNCS Volume 8591, 2014, pp. 453–464.
5 M. T. Omran, J.-R. Sack and H. Zarrabi-Zadeh, "Finding Paths with Minimum Shared Edges", (special issue of selected papers from COCOON 2011), J. Comb. Optim. 26(4):709–722, 2013
6 D. Nussbaum, M. T. Omran, J.-R. Sack, "Techniques to Protect Privacy Against Inference Attacks in Location Based Service", 3rd ACM SIGSPATIAL International Workshop on GeoStreaming (IWGS), 2012.

3.14 How does real-time information change the behaviour of traffic, and can cooperation emerge with intention sharing? (The online routing game model)

Laszlo Zsolt Varga (Eötvös Lorand University – Budapest, HU)

Active participants (agents) of urban transport are embedded in their environment. They perceive the current state of the traffic and make decisions which action to perform, e.g. which means of transportation to take, which route to follow, etc. If each subsequent agent reactively optimizes its trip, based on the real-time data of the current state of the whole traffic network, then in some situations the traffic may start to fluctuate, and sometimes the agents may be worse off with real-time data than without real-time data. This was observed by traffic engineers and multi-agent researchers both in real world and in simulations, and it was formally underpinned recently, using the novel online routing game model. Researchers proposed to add proactive flavour to the actions of the agents with anticipatory techniques, where the future state of the traffic is predicted from the current intentions of the agents.

Simulation results indicated, that these anticipatory techniques improve the properties of the traffic.

I am planning to present the online routing game model and how the model can be used to investigate these issues.

3.15 Spatio-temporal Search for Transportation Resources

Ouri E. Wolfson (University of Illinois – Chicago, US)

This talk will focus on the problem of spatio-temporal resource search and its application to transportation as in vehicular parking. Spatio-temporal resource search consists of two sub-problems: the location-detection of spatial resources, and the competition among mobile agents for capturing them. I will describe a crowdsourcing approach for the first, and a game theoretic approach for the second. I will discuss the implementation of the results in a new smartphone app that helps drivers locate open street-parking slots and capture them in competitive scenarios. In this sense, the app extends the car-navigation-system concept to parking.

4 Working groups

4.1 Benchmarking

Caitlin Doyle Cottrill (University of Aberdeen, GB), Ana Lucia Bazzan (Federal University of Rio Grande do Sul, BR), Andreas Hotho (Universität Würzburg, DE), Kevin Tierney (Universität Paderborn, DE), and Ronald Van Katwijk (TNO Telecom – Delft, NL)

In our first small group discussion, we reviewed the concept of benchmark data. In the discussion, it emerged that we all find that the concept of benchmark data is a good thing. Having a standard set of data that can be used for comparing processes and programs, perhaps particularly in the areas of model calibration and paper writing and reviewing, was viewed as a useful goal to pursue. It was also agreed, however, that establishing benchmark data is a complex process. Issues such as the following can negatively impact upon the potential for creating useful, standardized benchmark data sets:

- defining the standard the data needs to meet (good enough or gold standard),
- having access to adequate data (via sensors, surveys, etc.),
- defining the scenario(s) for which the data should correspond (i.e. timeliness, dispersion, context, etc.).

The ensuing discussion focused on both the potential for collecting benchmark data, as well as the purposes for which they could be used. Given the diversity of interests represented around the table (from freight to behaviour), it was quickly evident that no one data set could function in a similar fashion for all interests. There was discussion around the potential for creating (or identifying) a 'model city', which would be used for purposes of creating a 'perfect' synthetic or real dataset by instrumenting all aspects of infrastructure and behaviour. Ensuing discussion, however, revealed that this, too, would be in question. For instance, while it might be viewed as relatively simple to create a representative O-D matrix using a combination of in-road sensors, smartphones, and cameras, there would still likely be aspects (such as trajectory) lost in the process. Such discussions revealed both the interest in and relevance of the topic, while highlighting the myriad concerns evident in the construction and/or creation. The group broke up after this meeting; however, the underlying discussion and concerns were brought forward in ensuing discussions among the Dagstuhl participants.

4.2 Collaborative Travel in Public Transportation

Ivana Dusparic (University College Dublin, IE), Franziska Klügl (University of Örebro, SE), Marco Mamei (University of Modena, IT), Andrea Prati (University of Parma, IT), Jörg-Rüdiger Sack (Carleton University – Ottawa, CA), and Laszlo Zsolt Varga (Eötvös Lorand University – Budapest, HU)

Increasing proliferation of real-time travel information and travel planners as well as change in vehicle ownership rates are giving rise to novel transport models such as car-sharing, car-pooling, on-demand public transport, and provision of multi-modal transport as a service. One area that is still relatively unexplored is ad-hoc group travel in public transport, with the goal of obtaining significant group travel discounts transport companies offer to travellers, but also to provide more inclusive travel options to those who require additional help using the public transport (e.g., elderly, disabled, those who do not speak the local language etc). Transport providers would further benefit from such a model, by gaining new customers who would otherwise not be able to afford the travel, as well as from more fine-grained transport demand and preference information that groups formation would offer.

Self-organizing collaborative group formation for group public transport travel raises several challenges such as identification of suitable groups and group members based not only on constraints on source, destination and times of travel but also on group member preferences (e.g., which travel mode a user prefers, and what profile of group members does he or she prefer travelling with), appropriate group sizes sufficient to avail of discounts, re-organizing of the groups in case of cancellations, real-time delay updates (especially in the cases of connecting modes of transport), incorporating feedback on traveller reliability, and privacy of the travellers. In order to implement such a model, advances will need to be made in a number of research areas such as privacy, real-time demand prediction and simulation, learning of user preferences, and theoretical analysis of group formation. Impact of legal issues and regulations will also have to investigated and addressed.

Privacy: Questions on privacy falls into at least three different levels:
1. to the booking service,
2. to the assembled group and
3. to the travel provider.

Subscribers using such systems are willing to sacrifice some amount of privacy, in particular, when they are meeting face-to-face for travel in the assembled group (where/if joint travel is required). However, they may wish keep many aspects of their identity to themselves; these may include: exact locations of origin and destination and name (except for payment details and ticket verification). Obfuscation of exact locations, or techniques to ensure k privacy may be required while still allowing efficient group formation. These techniques need to be studied w.r.t. the requirements of the system and possibly new techniques may need to be developed in particular since privacy is here multi-level.

Learning of user preferences: Users might have multiple objectives and constraints on meeting those objectives when availing of public transport and group formation in particular. User preferences need to be captured, explicitly or implicitly, and their relative importance determined depending on the current context of the user. (e.g., users might prefer later train but are willing to change travel time within x minutes, for y amount of discount, but only if the weather is nice and they are not a personal and not business trip, and are alone rather than with a co-worker or a child). Implicit capturing of user preferences might include to access to sensory data, e.g. from the mobile phone of the user, while keeping in mind privacy issues. Suitable learning techniques will need to be identified and extended to deal with dynamically changing preferences and trade-offs between multiple objectives travel plans need to be optimized for.

We propose to perform a proof of concept study of such a system in public transport system (trains primarily) in Netherlands.

4.3 Online Data and Simulation

Franziska Klügl (University of Örebro, SE), Catherine Cleophas (RWTH Aachen, DE), Piyushimita Vonu Thakuriah (University of Glasgow, GB), and Laszlo Zsolt Varga (Eötvös Lorand University – Budapest, HU)

The title of the working leaves room for interpretation. During the seminar, we have seen that online web data is a valuable resource for many questions related to urban systems. For example, Twitter messages can be classified whether it contains a positive comment on public transportation or a comment about its safety. The term "prediction" is hereby used for predicting future classifications. Twitter messages like these could be used in a simulation model related to route choice. Data generated from online channels can be actually used for simulation like any other data source. The alternative to interpret the topic is the idea to feed data into a running simulation "online" – in contrast to offline, that means before the simulation or after the experiments for validation of simulation output. Thus approaches are also subsumed under the term dynamic data-driven simulation. Hereby, feeding-in data is associated with real-time data processing. Assuming that a simulation produces output y from input data x simulating a function like $y = a \cdot x + b$. Hereby, x is some form of input value, determines the output value of the simulation, a and b are parameters. We identified four approaches to feed in data during a running simulation.

1. The initial situation $x0$ may be directly connected to a real simulation. A simulation faster than real-time could illustrate how the world would develop starting from the current situation. An example would be the illustration of effects of climate change.
2. Online infusion of input data. There is an ongoing connection between real-world sensors and a continually running simulation. Sensor events are fed into the simulation. An

example is the OLSIM Simulation which connects a simulation-based extrapolation of traffic state of the highway network of North-Rhine Westphalia (Germany) to sensors that register vehicles moving on and off ramps.
3. Online calibration would address adaptation of parameters such as a or b so that the current simulation is better aligned to the real world.
4. Model adaptation means a data-driven alignment of a running simulation to the real-world by not just adapting parameter, but changing the structure of the model. In the example that means the way how the input values and parameters are combined, the $*$ or $+$.

Whether such a dynamic data-driven approach could work depends on the temporal resolution of model dynamics with respect to granularity of incoming sensor data and the time frame of predictions that the model could provide:

- Short time appears to be quite clearly related: the simulation could provide forecasts of travel time given the current real-world traffic state.
- Mid-term would relate to daily activities, mode choice models; Real people's choices can be clearly observed and extrapolated into the future.
- Longer-terms like simulations for finding the optimal headway of busses or adaptation of infrastructure appear to be harder.

Based on that discussion, we identified interesting application ideas:

- Simulating the urban metabolism consisting of different sub-systems with different data needs aligned to the real world. One could identify locations at which there might be a big crowd in an hour. Based on those predictions – if done reliably – actions could be triggered such as send more busses than originally planned.
- Knowing that a fitness app tells a person to drop of the bus now for satisfying the daily goals of number of steps, could be fed into the prediction of travel time for the person that drops off, but also for the complete system.
- Predicting whether future connections would work given number of travelers who intend to go to destination and would due to their numbers further delay the connections at other places.
- Simulation for predicting my information needs about an activity before I actually fully planned the activity. As a result of this discussions, we decided to continue with personalized travel information also involving simulation-based predictions.

4.4 hello... Situation Aware Systems Integration (SASI)

Thomas Leo McCluskey (University of Huddersfield, GB), Sybil Derrible (University of Illinois – Chicago, US), Cecilia Gomes (New University of Lisbon, PT), Jörn Schlingensiepen (TH Ingolstadt, DE), Piyushimita Vonu Thakuriah (University of Glasgow, GB), and Ronald Van Katwijk (TNO Telecom – Delft, NL)

Cities have gained tremendous complexity since the industrial revolution, forcing the separation of responsibilities across multiple "departments" (e.g. transport, water resources, utilities). This directly contrasts with older paradigms of planning that were led by central authorities (Figure 1), what Lewis Mumford (1961, p. 172) describes as "regimenters of human functions and urban space" and perhaps best illustrated by Baron Haussmann for

Paris and Ildefons Cerdà for Barcelona. Beyond the substantial benefits, this decentralization has come at a cost, and nowadays, departments rarely communicate with one another, acting as distinct silos. What is more, the situation is likely to be exacerbated by the predicted rise in urban population through this century. Letting departments develop vertically leads to interference of services (perhaps discovered in operation), lack of coordination when facing urban-wide challenges and emergencies, and inherent inefficiencies in the whole urban system. While there has been work in service / department co-ordination for example in the area of disaster recovery, by and large these developments tend to be ad hoc.

While many argue for a breaking down of the silos, there are two ways that this can occur. One way is to re-centralize responsibilities. This is unlikely, however, since cities are complex systems, and it is not reasonable to assume that one entity can successfully plan and coordinate all activities of a city. A more preferable paradigm is to encourage communication between departments, as illustrated on the right-hand-side of Figure 2. To this end, the main goal of this project is to enhance department-to-department (D2D) cooperation and coordination, via improved D2D communication and data exchange, through a common formal semantic vocabulary. This is particularly useful for data acquisition purposes. For instance, the transportation department will be able to communicate with other departments and share services and data for the benefit of the city; e.g., the transportation department will easily be able to acquire urban emission data from the environmental department. To some extent "smart cities" initiatives are aimed at such improvements, but up to now the focus of smart cities has been on sharing data through enhanced ICT infrastructure. Little work has focused on investigating the requirements on and functions of departments (utilities) with the aim of considering these together with the aims of coordination, cooperation and synchronization.

To illustrate the problem, we detail two "examples" of how a city with isolated functions/ utilities might fail: (1) urban emissions, and (2) urban flooding.

Example 1: Transport, energy generation and construction are three "departments" within a City that manage functions which generate urban emissions as a side effect. Currently there is no (as far we know) communication or co-ordinating control between them. It is well known that cities regularly break air pollution limits, even in Europe. It may be possible for "Transport" to adjust regional traffic flows to alleviate this pollution, but this takes no account of the other ca. 50% of pollution which is generated by other utilities. Currently it is not known how these departments can communicate and combine together to produce a holistic solution to the problem. In fact, there is likely to be little shared knowledge between them. Enabling D2D communication on a shared vocabulary, could significantly tackle this challenge.

Example 2: Flooding has become a significant threat in cities. The impervious surfaces created by buildings and roads force storm water systems to handle large flows of surface runoffs during heavy rains. Runoffs access storm water sewers through grates at street curbs that can easily be obstructed by debris, causing flooding. This problem is likely to be amplified in the future as extreme weather events are not only increasing in frequency but also in number, thus making cities particularly vulnerable to flooding. At the moment, municipal crews drive around neighborhoods before and during rain events, and try to detect obstructed grates. This information could easily be collected by other parties, however, from being crowd-sourced by responsible citizens or police patrols, to being detected by smart streetlights (that can detect when flooding when equipped with sensors). Although water departments could implement their own monitoring system (e.g., in the form of 311 calls such as in Chicago)), the goal here is to offer a common language and make it easy to share information across departments, beyond implementing an ad hoc system that purely fits this particular case. A similar process can be imagined for many other problems, from pothole detection to sag in distribution lines, which could easily be reported by non-responsible parties.

Methodology

Since the interaction of different stakeholders in this field is too complex to drive this development in a top-down-way, we propose to uses an approach inspired by functional design. The main idea is to describe current and future demands in a solution-independent way in order to free the mind in order to think about the solution in new ways. The main aspect is to give a value to provide functions but also allow identifying advantages in coupling different functions / services in cities. We aim to capture the requirements of the example systems in a formal way, and hence capturing the semantics of the functions and services, and the semantics of the data flowing between them, independent of our interpretation. Data sets could be annotated with ontologies which will underpin the sharing of data. Additionally, this will enable the use of tools to look for inconsistencies and interference between requirements. (For example, we would be able to check whether, with the addition of some services requirements, safety conditions on some existing service are compromised). More generally, it would allow us to simulate the behavior of these systems defined by the requirements. It will also facilitate the modular addition of the requirements of further services in the future. The expression of these requirement will rely on the use of a fundamental commitment to some common language convention, such as first order logic, using a controlled but extensible vocabulary.

Since data acquisition is a big cost factor in providing smart services the focus is on identifying data that can be shared or data that is demanded by one service solution that is already generated by another. To achieve this we propose to focus on coupling two or three current services resp. their providers in the first step.

Intersection (re-)design

Intersections are locations where many people meet that are in transit. In this case we are focusing on locations where different modes of transport, often with different speeds, meet. Currently the merits of one solution over the other is made by quantifying only one or two criteria, which are mostly internal to the transport system. Solutions also affect the city in other important ways, which are at best taken into account only qualitatively.

Key performance indicators:

- Safety (both objective (crashes) and subjective (social safety),
- Fairness (with respect to waiting times),
- Efficiency (with respect to energy use, throughput),
- Aesthetically pleasing
- Attractiveness of nearby housing and shops
- Healthy (emissions and emissions)

Known (partial) solutions

Shared space, spatial segregation of different travel modes (vertically by means of underpasses and overpasses or horizontally by providing separate infrastructure), signalization, roundabouts, speed harmonization

Data requirements:

- Expected) traffic demand for the intersection split out for
- the different modes of transport: motorized traffic, bicycles (electric and manual), pedestrians,
- The different (types) of users of the intersection (elderly, schoolchildren, shoppers, etc.)
- How demand could be affected by the chosen solution both locally and elsewhere in the network
- The available resources (time, space and money)
- Number of residents and merchants effected economically or health-wise (emissions and emmisions)

Some informative sources:

- http://www.eltis.org/guidelines/introduction
- https://en.wikipedia.org/wiki/Shared_space
- http://www.pps.org/reference/shared-space/
- https://mitpress.mit.edu/sites/default/files/titles/content/9780262012195_sch_0001.pdf
- http://nacto.org/publication/urban-street-design-guide/intersections/intersection-design-principles/

Contribution of Maria Cecilia

The idea of Cloud Computing was preceded by the idea of Grid computing (similar to the power Grid), i.e. having computational power everywhere, you just have to "plug your computer". Since big enterprises like Amazon and Google had big data centers, they started offering storage and computational resources as a service., leading to the basics of Cloud computing – IaaS (Infrastructure as a Service) PaaS (Platform as a S), SaaS (Software as a Service). So the Cloud computing is already making the cities smarter, and helping transportation by avoiding transportation, but mostly by providing resources in a transparent and distributed way, as with a cost model "pay as you need".

Cloud computing is hence providing services for:

- people
- commercial enterprises
- transport domain (private and public)
- authorities (e-city)
- entertainment enterprises

References

1 L. Mumford. *The city in history: its origins, its transformations, and its prospects*. Harcourt, Brace & World, New York, 1961.

4.5 COllaborative Flexible Freight for Eco-Efficient Delivery (COFFEE-D)

Kevin Tierney (Universität Paderborn, DE), Catherine Cleophas (RWTH Aachen, DE), Caitlin Doyle Cottrill (University of Aberdeen, GB), Jan Fabian Ehmke (Freie Universität Berlin, DE), and Benjamin Kickhöfer (TU Berlin, DE)

This group addressed the topic of collaborative freight transportation in urban environments. Urbanization has been steadily increasing since the industrial revolution and sees no signs of abating. As more and more people relocate to cities, the challenges for urban governments and authorities are becoming larger and more difficult to overcome. Pollution, noise, and traffic congestion are several key problems in urban areas requiring action to avoid negative consequences for the health and well being of city inhabitants. With residents demanding a higher quality of life in their neighborhoods, innovative solutions are required to address the downsides of city growth.

As cities expand, so does the amount of freight shipped in to, out of and within urban areas. Freight carrying trucks are key contributors to lower air quality, noise, and congestion of streets both in terms of parking and traffic flow. However, truck traffic in cities is unavoidable according to current freight transportation paradigms. Large cities consume and produce many tons of goods and require constant traffic in and out of the city. An innovative transportation solution would therefore reduce truck traffic within cities while continuing to support the freight demands of a city's inhabitants.

We propose utilizing existing light rail infrastructure to transport goods from transshipment hubs located outside of a city to stations within the city. Receivers of goods would either pick up their goods in package stations located along light rail lines or would engage third party last mile services to bring packages to their doorstep. The system would support shipments from and to anybody inside of a city, meaning unlike previous light rail-based freight systems, it would also be able to offer last mile services normally provided by the post or package companies like DHL, UPS or Fed-Ex. Since the entire system would run on light rail, it would be clean and not detract from the city's air quality. And it would also run independently from road traffic.

The COFFEE-D system would use a set of stackable containers in various sizes that can be palletized as well as carried on hand trucks or by a single person. Goods destined for locations within the city would be placed into these containers at their place of production or packing, loaded into trucks and brought to transshipment hubs at the border of a city. Automated systems would quickly load the containers into a buffer at the hub and await a freight pod to carry containers into the city. Freight pods could be fully automated, semi-automated, or towed and would directly interface with package stations at selected stops.

The outcome of our proposed system will be cleaner cities with less traffic and parking congestion. We plan to submit a grant in a Horizon 2020 EU call to further explore the COFFEE-D system.

4.6 Modelling Pedestrian Behaviour

Sabine Timpf (Universität Augsburg, DE), Benjamin Heydecker (University College London, GB), Andreas Hotho (Universität Würzburg, DE), Tobias Kretz (PTV AG – Karlsruhe, DE), Daniele Quercia (NOKIA Bell Labs – Cambridge, GB), and Giuseppe Vizzari (University of Milano-Bicocca, IT)

The reason to model pedestrian behaviour within the context of CTS is to gain an understanding of this behaviour and to test hypotheses on behaviour. One motivation for this modeling might be to inform the design of pedestrian systems; this is especially true in event planning where issues for investigation need to be identified. Offline simulations and analysis of results may improve key performance indicators (Level Of Service, travelling times, densities, etc.); faster-than-realtime simulations may be used for short term predictions (e.g. in systems supporting crowd managers / operators) or to inform individual pedestrians as a basis for a decision-support system or for a cooperative system. Modelling pedestrian behaviour is a multi-perspective endeavour, which means that planners, architects, urban and landscape designers, transportation engineers and managers need to bring their own issues and questions in order for a satisfactory model to emerge. Bringing together this many disciplines remains a challenge even through the intermediary of models. In a modelling framework many influences on pedestrian behaviour need to be considered: in general the spatial cognition of pedestrians, their situational awareness, their objective(s), their knowledge of routes and areas, and their culture. Within the model run the objectives of activity need to be taken into account, the information levels, the development of knowledge of time, the proximity and the physical contact with others need to be modelled. Apart from these parameters we are very much interested in defining and deriving the influence of cooperation among pedestrians.

Cooperative Pedestrians

Cooperation can happen at several levels, with increasing prospective benefits to participants individually and the community as a whole. Cooperation might be positive and voluntary, but also forced (i.e. through social norms or rules) or triggered by incentives. Cooperation is founded on communication of data and information of various kinds. One possible classification of levels of communication, with increasing prospects for cooperation, is as follows:

- Sharing data (or information) about current state (position, speed, instantaneous properties)
- Sharing knowledge of individual decisions that have been made, based on individual information and information from others
- Sharing intentions (e.g., intended destination - both location and schedule)
- Making joint decisions based on shared information and knowledge of intentions.This seems to be collaboration, which goes beyond cooperation

First try at making a definition

Cooperative Pedestrians share the same space but they generally have different goals leading to potentially conflicting use of the shared space. Cooperative behavior arises when individual actions are not optimal (from a personal perspective) but improve the general welfare (e.g. improve the overall flow). The individual tendency for cooperative behavior must be there or must be enforced. The basis for cooperative behavior seems to be the sharing of information with resulting shared intentions/objectives.

Examples of ways of cooperating

1. Pedestrians cooperating among themselves:
   - avoid moving between perceived members of a group
   - groups can be:
     - 2 peers (friends, partners)
     - parent+child or similar hierarchical pair with clearly defined leader
     - group with a leader (think tourists with a guide)
     - unguided group of peers (consider a party of friends)
   - changing behavior to account for density, i.e. forming lanes with your own people
   - cooperation in panic situations, people also help each other (we're better off if we stick together), research on influence of expectation of behavior needed
2. Ped cooperating with management system, taking escalator up, but walking down stairs
3. Ped sharing the same goal
4. Ped coming towards each other
5. Ped standing on right and walking on left on escalator

The altruistic pedestrian who, finding no space on the right where they wish to stand, walks up on the left when they do not wish to walk but does so for the benefit of those walking up on the left behind them.

Examples of (possibly) cooperating pedestrians

1. Alighting and boarding tube train
2. Themepark with different queues
3. Pop-concert with different stages
4. Marathon runners turning the corner
5. Fans approaching a stadium
6. Cordon of cleaners coordinating to also clear an arena of people (walking in formation)

Models

Among the different approaches to modeling pedestrian behaviour there is a need to compare microsimulation models with agent-based simulation and with physics-based models. It is not clear how to setup such a comparison or how to validate these models. Measures for validation are also needed.

Data

This type of research requires empirical evidences characterizing pedestrian behavior: this kind of information may be gathered by means of controlled experiments or observations under natural conditions. Both approaches have advantages and issues, for instance the possibility to generalize the acquired results in different contexts (e.g. cultural differences, motivations of participants or observed pedestrians), so generally both are necessary. As of this moment, although there is a growing interest in both large scale experiments and observations, but there are no guidelines on how to conduct these activities. In case of analyses carried out through video recording, post processing should employ state of the art or innovative computer vision tools to reduce the necessary effort. Moreover, although there are efforts towards the definition of a shared standard way to validate simulation models (e.g.

NIST Technical Note 1822), also this kind of activity could benefit from additional reflection, possibly involving different disciplinary perspectives.

Different pedestrians move differently and one and the same pedestrian may move differently at different times. This roots in the motivation for movement. Motivation in turn depends on external properties (e.g. air temperature) and internal properties which can be goal oriented (hurry to reach a store before it closes for the day) or internally enforced (keep different distance to different types of others as a consequence of social norms). All this implies that a seemingly simple property like "free speed" in fact already is a difficult theoretical concept and subject to temporal variations and influences which are not necessarily reflected by models. To come up with proper models one needs to analyze not only the observable properties like speed or trajectories but also the more general goals and preferences which can be gathered by controlled experiments.

Dimensions of analysis

Among the dimensions of analysis in transport, we are interested in the following: The first dimension concerns cargo versus persons: we are interested only in persons. The second dimension deals with system design versus operation support. We believe that pedestrian behavior models will be of benefit to both in the form of offline simulations which support design and planning, whereas the gathering and sharing of data (maybe with some elaborations) supports operations. Within the third dimension (individual vs many) we are interested in the many, first, because they produce data, but they also are provided with awareness information to support their decisions. The move from individual decision to collaborative decision is not clear and needs further investigation. As for the fourth dimension (centralized vs decentralized control); the term control is probably inappropriate: the system provides information that can support cooperative behaviors, however, "control" is decentralized. This also need further research. Within the dimension descriptive vs prescriptive our main focus lies on descriptive modelling, but this can ultimately be used to generate prescriptive advice to pedestrians, to managers and to designers.

Participants

- Niels Agatz
Erasmus University – Rotterdam, NL
- Ana Lucia Bazzan
Federal University of Rio Grande do Sul, BR
- Catherine Cleophas
RWTH Aachen, DE
- Caitlin Doyle Cottrill
University of Aberdeen, GB
- Sybil Derrible
University of Illinois – Chicago, US
- Ivana Dusparic
University College Dublin, IE
- Jan Fabian Ehmke
FU Berlin, DE
- Cecilia Gomes
New University of Lisbon, PT
- Benjamin Heydecker
University College London, GB
- Andreas Hotho
Universität Würzburg, DE
- Benjamin Kickhöfer
TU Berlin, DE
- Franziska Klügl
University of Örebro, SE
- Tobias Kretz
PTV AG – Karlsruhe, DE
- Ronny Kutadinata
The University of Melbourne, AU
- Marco Mamei
University of Modena, IT
- Dirk Christian Mattfeld
TU Braunschweig, DE
- Thomas Leo McCluskey
University of Huddersfield, GB
- Andrea Prati
University of Parma, IT
- Daniele Quercia
NOKIA Bell Labs – Cambridge, GB
- Jörg-Rüdiger Sack
Carleton University – Ottawa, CA
- Jörn Schlingensiepen
TH Ingolstadt, DE
- Monika Sester
Leibniz Univ. Hannover, DE
- Piyushimita Vonu Thakuriah
University of Glasgow, GB
- Kevin Tierney
Universität Paderborn, DE
- Sabine Timpf
Universität Augsburg, DE
- Eric van Berkum
University of Twente, NL
- Ronald Van Katwijk
TNO Telecom – Delft, NL
- Laszlo Zsolt Varga
Eötvös Lorand University – Budapest, HU
- Giuseppe Vizzari
University of Milano-Bicocca, IT
- Ouri E. Wolfson
University of Illinois – Chicago, US

Report of Dagstuhl Seminar 16092

Computational Music Structure Analysis

Edited by
Meinard Müller[1], Elaine Chew[2], and Juan Pablo Bello[3]

1 **Universität Erlangen-Nürnberg, DE,** `meinard.mueller@audiolabs-erlangen.de`
2 **Queen Mary University of London, GB,** `elaine.chew@qmul.ac.uk`
3 **New York University, US,** `jpbello@nyu.edu`

Abstract

Music is a ubiquitous and vital part of the lives of billions of people worldwide. Musical creations and performances are among the most complex and intricate of our cultural artifacts, and the emotional power of music can touch us in surprising and profound ways. In view of the rapid and sustained growth of digital music sharing and distribution, the development of computational methods to help users find and organize music information has become an important field of research in both industry and academia.

The Dagstuhl Seminar 16092 was devoted to a research area known as music structure analysis, where the general objective is to uncover patterns and relationships that govern the organization of notes, events, and sounds in music. Gathering researchers from different fields, we critically reviewed the state of the art for computational approaches to music structure analysis in order to identify the main limitations of existing methodologies. This triggered interdisciplinary discussions that leveraged insights from fields as disparate as psychology, music theory, composition, signal processing, machine learning, and information sciences to address the specific challenges of understanding structural information in music. Finally, we explored novel applications of these technologies in music and multimedia retrieval, content creation, musicology, education, and human-computer interaction.

In this report, we give an overview of the various contributions and results of the seminar. We start with an executive summary, which describes the main topics, goals, and group activities. Then, we present a list of abstracts giving a more detailed overview of the participants' contributions as well as of the ideas and results discussed in the group meetings of our seminar.

Seminar February 28–March 4, 2016 – http://www.dagstuhl.de/16092
1998 ACM Subject Classification H.5.5 Sound and Music Computing
Keywords and phrases Music Information Retrieval, Music Processing, Music Perception and Cognition, Music Composition and Performance, Knowledge Representation, User Interaction and Interfaces, Audio Signal Processing, Machine Learning
Digital Object Identifier 10.4230/DagRep.6.2.147
Edited in cooperation with Stefan Balke

1 Executive Summary

Meinard Müller
Elaine Chew
Juan Pablo Bello

In this executive summary, we start with a short introduction to computational music structure analysis and then summarize the main topics and questions raised in this seminar.

Furthermore, we briefly describe the background of the seminar's participants, the various activities, and the overall organization. Finally, we reflect on the most important aspects of this seminar and conclude with future implications and acknowledgments.

Introduction

One of the attributes distinguishing music from other types of multimedia data and general sound sources are the rich, intricate, and hierarchical structures inherently organizing notated and performed music. On the lowest level, one may have sound events such as individual notes, which are characterized by the way they sound, i.e., their timbre, pitch and duration. Such events form larger structures such as motives, phrases, and chords, and these elements again form larger constructs that determine the overall layout of the composition. This higher structural level is specified in terms of musical parts and their mutual relations. The general goal of *music structure analysis* is to segment or decompose music into patterns or units that possess some semantic relevance and then to group these units into musically meaningful categories.

While humans often have an intuitive understanding of musical patterns and their relations, it is generally hard to explicitly describe, quantify, and capture musical structures. Because of different organizing principles and the existence of temporal hierarchies, musical structures can be highly complex and ambiguous. First of all, a temporal segmentation of a musical work may be based on various properties such as homogeneity, repetition, and novelty. While the musical structure of one piece of music may be explained by repeating melodies, the structure in other pieces may be characterized by a certain instrumentation or tempo. Then, one has to account for different musical dimensions, such as melody, harmony, rhythm, or timbre. For example, in Beethoven's Fifth Symphony the "fate motive" is repeated in various ways – sometimes the motive is shifted in pitch, sometimes only the rhythmic pattern is preserved. Furthermore, the segmentation and structure will depend on the musical context to be considered; in particular, the threshold of similarity may change depending on the timescale or hierarchical level of focus. For example, the recapitulation of a sonata may be considered a kind of repetition of the exposition on a coarse temporal level even though there may be significant modifications in melody and harmony. In addition, the complexity of the problem can depend on how the music is represented. For example, while it is often easy to detect certain structures such as repeating melodies in symbolic music data, it is often much harder to automatically identify such structures in audio representations. Finally, certain structures may emerge only in the aural communication of music. For example, grouping structures may be imposed by accent patterns introduced in performance. Hence, such structures are the result of a creative or cognitive process of the performer or listener rather then being an objective, measurable property of the underlying notes of the music.

Main Topics and Questions

In this seminar, we brought together experts from diverse fields including psychology, music theory, composition, computer science, music technology, and engineering. Through the resulting interdisciplinary discussions, we aimed to better understand the structures that emerge in composition, performance, and listening, and how these structures interrelate. For example, while there are certain structures inherent in the note content of music, the perception and communication of structure are themselves also creative acts subject to interpretation. There may be some structures intended by the composer or improviser, which are not fully communicated by symbolic descriptions such as musical score notation. The performer, if different from the composer, then must interpret structures from the score, and decide on the prosodic means by which to convey them. When a listener then tries to make sense of the performed piece, that act of sense-making, of constructing structure and meaning from an auditory stream is also a creative one. As a result, different people along this communication chain may come up with different solutions, depending on their experiences, their musical backgrounds, and their current thinking or mood.

Based on our discussions of various principles and aspects that are relevant for defining musical patterns and structures, the following questions were raised.

- How can ambiguity in notions such as repetition, similarity, grouping, and segmentation be handled and modeled?
- In which way do these notions depend on the music style and tradition?
- How can one account for the relations within and across different hierarchical levels of structural patterns?
- How can long-term structures be built up from short-term patterns, and, vice versa, how can the knowledge of global structural information support the analysis of local events?
- How can information on rhythm, melody, harmony, timbre, or dynamics be fused within unifying structural models?
- How can the relevance of these aspects be measured?
- How do computational models need to be changed to account for human listeners?

By addressing such fundamental questions, we aimed for a better understanding of the principles and model assumptions on which current computational procedures are based, as well as the identification of the main challenges ahead.

Another important goal of this seminar was to discuss how computational structure analysis methods may open up novel ways for users to find and access music information in large, unstructured, and distributed multimedia collections. Computational music structure analysis is not just an end in itself; it forms the foundation for many music processing and retrieval applications. Computational methods for structuring and decomposing digitized artifacts into semantically meaningful units are of fundamental importance not only for music content but also for general multimedia content including speech, image, video, and geometric data. Decomposing a complex object into smaller units often constitutes the first step for simplifying subsequent processing and analysis tasks, for deriving compact object descriptions that can be efficiently stored and transmitted, and for opening up novel ways for users to access, search, navigate, and interact with the content. In the music context, many of the current commercially available services for music recommendation and playlist generation employ *context-based* methods, where textual information (e. g., tags, structured metadata, user access patterns) surrounding the music object are exploited. However, there are numerous data mining problems for which context-based analysis is insufficient, as it tends to be low on specifics and unevenly distributed across artists and styles. In such cases, one requires *content-based* methods, where the information is obtained directly from the analysis of audio signals, scores and other representations of the music. In this context, the following questions were raised.

- How can one represent partial and complex similarity relations within and across music documents?
- What are suitable interfaces that allow users to browse, interact, adapt, and understand musical structures?
- How can musical structures be visualized?
- How can structural information help improve the organizing and indexing of music collections?

Participants, Interaction, Activities

In our seminar, we had 31 participants, who came from various locations around the world including North America (8 participants from the U.S.), Asia (2 participants from Japan), and Europe (21 participants from Austria, France, Germany, Netherlands, Portugal, Spain, United Kingdom). Many of the participants came to Dagstuhl for the first time and expressed enthusiasm about the open and retreat-like atmosphere. Besides its international character, the seminar was also highly interdisciplinary. While most of the participating researchers are working in the fields of music information retrieval, we have had participants with a background in musicology, cognition, psychology, signal processing, and other fields. This led to the seminar having many cross-disciplinary intersections and provoking discussions as well as numerous social activities including playing music together. One particular highlight of such social activities was a concert on Thursday evening, where various participant-based ensembles performed a wide variety of music including popular music, jazz, and classical music. Some of the performed pieces were original compositions by the seminar's participants.

Overall Organization and Schedule

Dagstuhl seminars are known for having a high degree of flexibility and interactivity, which allows participants to discuss ideas and to raise questions rather than to present research results. Following this tradition, we fixed the schedule during the seminar asking for spontaneous contributions with future-oriented content, thus avoiding a conference-like atmosphere, where the focus tends to be on past research achievements. After the organizers have given an overview of the Dagstuhl concept and the seminar's overall topic, we started the first day with self-introductions, where all participants introduced themselves and expressed their expectations and wishes for the seminar. We then continued with a small number of ten-minute stimulus talks, where specific participants were asked to address some critical questions on music structure analysis in a nontechnical fashion. Each of these talks seamlessly moved towards an open discussion among all participants, where the respective presenters took over the role of a moderator. These discussions were well received and often lasted for more than half an hour. The first day closed with a brainstorming session on central topics covering the participants' interests while shaping the overall schedule and format of our seminar. During the next days, we split into small groups, each group discussing a more specific topic in greater depth. The results and conclusions of these parallel group sessions, which lasted between 60 to 90 minutes, were then presented to, and discussed with, the plenum. Furthermore, group discussions were interleaved with additional stimulus talks spontaneously given by participants. This mixture of presentation elements gave all participants the opportunity for presenting their ideas to the plenum while avoiding a monotonous conference-like presentation format. Finally, on the last day, the seminar concluded with a session we called "self-outroductions" where each participant presented his or her personal view of the main research challenges and the seminar.

Conclusions and Acknowledgment

Having the Dagstuhl seminar, our aim was to gather researchers from different fields including information retrieval, signal processing, musicology and psychology. This allowed us to approach the problem of music structure analysis by looking at a broad spectrum of data analysis techniques (including signal processing, machine learning, probabilistic models, user studies), by considering different domains (including text, symbolic, image, audio representations), and by drawing inspiration from creative perspectives of the agents (composer, performer, listener) involved. As a key result of this seminar, we achieved some significant progress towards understanding, modeling, representing, extracting, and exploiting musical structures. In particular, our seminar contributed to further closing the gap between music theory, cognition, and the computational sciences.

The Dagstuhl seminar gave us the opportunity for having interdisciplinary discussions in an inspiring and retreat-like atmosphere. The generation of novel, technically oriented scientific contributions was not the focus of the seminar. Naturally, many of the contributions and discussions were on a rather abstract level, laying the foundations for future projects and collaborations. Thus, the main impact of the seminar is likely to take place in the medium to long term. Some more immediate results, such as plans to share research data and software, also arose from the discussions. As measurable outputs from the seminar, we expect to see several joint papers and applications for funding.

Beside the scientific aspect, the social aspect of our seminar was just as important. We had an interdisciplinary, international, and very interactive group of researchers, consisting of leaders and future leaders in our field. Many of our participants were visiting Dagstuhl for the first time and enthusiastically praised the open and inspiring setting. The group dynamics were excellent with many personal exchanges and common activities. Some scientists expressed their appreciation for having the opportunity for prolonged discussions with researchers from neighboring research fields – some thing that which is often impossible during conference-like events.

In conclusion, our expectations of the seminar were not only met but exceeded, in particular with respect to networking and community building. We would like to express our gratitude to the Dagstuhl board for giving us the opportunity to organize this seminar, the Dagstuhl office for their exceptional support in the organization process, and the entire Dagstuhl staff for their excellent service during the seminar. In particular, we want to thank Susanne Bach-Bernhard, Roswitha Bardohl, Marc Herbstritt, and Sascha Daeges for their assistance during the preparation and organizing of the seminar.

2 Table of Contents

3 Stimulus Talks

3.1 Computational Music Structure Analysis (or How to Represent the Group's Interests and Opinions in a Few Slides)

Juan Pablo Bello (New York University, US)

In this talk, I have tried to identify areas of commonality and divergence in the abstracts submitted before the seminar, with the goal of stimulating and seeding the discussions during the Dagstuhl seminar.

First, I highlighted the wide range of applications of interest to the participants, including those driven by musicological concerns like the analysis of Jazz improvisations, performances of Wagner operas, metrical structures in Carnatic music, and notions of flow in Rap music; music information retrieval applications such as automatic rhythm transcription from recorded music, optical music recognition and understanding the structure of large music collections to characterize patterns of originality; and more creative applications in augmenting live music performances, algorithmic composition, the automatic creation of mashups and remixes, and improving the workflow of music production. Paraphrasing Serra's point, it is critical to understand the differences between the signal's properties, the application requirements, and the user context, and design solutions accordingly.

And yet, the abstracts highlighted many common issues that cut across applications. For example, the complex relational structure of music including multiple, hierarchical levels of information with strong interdependencies, and the multi-faceted nature of music information. Also the difficulty of defining notions of similarity and dissimilarity governing those relationships, and of creating representations of music information that are either invariant or sensitive to these patterns of similarity. It is generally agreed upon that current computational methods fail to capture this complexity.

A number of methods and ideas were proposed for addressing this: Markov logic networks, graph analysis, uncertainty functions, combinatorial optimization, statistical regularities in annotations, the use of multiple and rich streams of information, the use of domain-knowledge or the formulation of structural analysis as an information theoretical problem. More fundamentally, the abstracts emphasized the need to revise core assumptions in past literature, notably the shortcomings of the widespread ground-truth paradigm in the context of a task that is ambiguous and thus lends itself to multiple interpretations and disagreements. In this scenario, what should be the goal of computational approaches? To return all or multiple interpretations? Or at least one interpretation deemed to be coherent (or reasonable, plausible, interesting)? How do we define or benchmark coherence? Is this connected to narrative, flow, grammatical consistency? These are open questions that the community needs to address to move the field forward.

Finally, there is the question of the role of humans in the different stages of this process: data collection, computational modeling and benchmarking, and whether the computational task should be redefined as a way to gain insight on the human processing of musical structures, and even the modeling of individual responses.

3.2 What You Hear and What You Must Make Others Hear

Elaine Chew (Queen Mary University of London, GB)

URL http://elainechew-research.blogspot.com/2016/03/new-thoughts-on-piano-performance-sound.html

Drawing is not form, but a way of seeing form.
Drawing is not what you see, but what you must make others see.
Drawing is not form, it is your understanding of form.
– Edgar Degas (Gammell 1961, p.22)

Music, and the performer or composer, works in much the same way. Alternative interpretations or hearings of musical form and structure almost always exist. These differences can be attributed to the listeners' state of knowledge, prior expectations, attention, and ontological commitment [1]. Given a particular hearing of musical structure, the performer can project this structure in performance through musical prosody [2]. The prosody of a musical communication strongly influences the listeners' parsing of its structure and meaning [3, 4]. In this talk, I gave numerous examples to show that structure emerges from performances, and that performances serve as a rich and largely untapped source of information about musical structure.

References

1 Smith, J. B. L., I. Schankler, E. Chew (2014). Listening as a Creative Act: Meaningful Differences in Structural Annotations in Improvised Performances. Music Theory Online, 20(3). http://www.mtosmt.org/issues/mto.14.20.3/mto.14.20.3.smith_schankler_chew.html
2 Chew, E. (2012). About Time: Strategies of Performance Revealed in Graphs. Visions of Research in Music Education 20(1). http://www-usr.rider.edu/~vrme/v20n1/visions/Chew%20Bamberger%20.pdf
3 Chew, E. (2016). Playing with the Edge: Tipping Points and the Role of Tonality. Music Perception, 33(3): 344-366. http://mp.ucpress.edu/content/33/3/344
4 Chew, E. (2016). From Sound to Structure: Synchronizing Prosodic and Structural Information to Reveal the Thinking Behind Performance Decisions. In C. Mackie (ed.): New Thoughts on Piano Performance. http://elainechew-research.blogspot.com/2016/03/new-thoughts-on-piano-performance-sound.html

3.3 Exposing Hierarchy Through Graph Analysis

Brian McFee (New York University, US)

Joint work of Juan Bello, Oriol Nieto, Dan Ellis, Brian McFee

In this talk, I presented a graphical perspective on musical structure analysis. While it is common to treat the boundary detection and segment labeling problems independently (or at least sequentially), viewing structure analysis as a graph partitioning problem can provide a unified formalization of both tasks, and motivate new algorithmic techniques. By varying the number of elements in the desired partition, it is possible to reveal multi-level or hierarchical structure in music. I concluded with a discussion of current challenges in both algorithm design and evaluation for hierarchical structure analysis.

References

1 Brian McFee, Oriol Nieto, and Juan Pablo Bello: Hierarchical Evaluation of Segment Boundary Detection. In Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR), Málaga, Spain, 406–412 , 2015.

2 Brian McFee and Dan Ellis: Analyzing Song Structure with Spectral Clustering. In Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan, 405–410, 2014.

3 Brian McFee and Daniel P. W. Ellis: Learning to segment songs with ordinal linear discriminant analysis, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 5197–5201, 2014.

3.4 Music Structure: Seeking Segmentations or Scenes?

Cynthia C. S. Liem (TU Delft, NL)

Joint work of Mark S. Melenhorst, Martha Larson, Alan Hanjalic, Cynthia C. S. Liem

When I was asked to give a Stimulus Talk at this Dagstuhl seminar, I reflected on my own past involvement with music structure analysis. As for me, this mostly fell into two categories: first of all, homework assignments for music theory courses back in conservatoire. Secondly, annotations I made for outcome visualizations of past research, in which developments over the course of a music performance (e.g. timing [5] and movement [2]) would partially relate to structure.

However, while I have always read work on automated music structure analysis in the community with interest, somehow I never felt the urge to work on the problem myself. Was it something I just took for granted, causing an interest mismatch similar to those collected within the community after a past Dagstuhl seminar [4]? I came to realize my own interests were not so much in localizing exact segment boundaries, but rather in the events happening in between such boundaries – and their contextualization with respect to various interpretations, at the performer and audience side.

In my talk, I reflected a bit more on this, discussing how notions of structure (even if ambiguous) can be a means to tackling higher-level, bigger questions on how music is realized and interpreted. I did this by discussing three topics:

- the role of structure in musical interpretation [5];
- the importance of narrative and linguistic event structure in music, when moving towards the connection of music to other media [3];
- the role of structure in assisting concert experiences of music audiences through digital interfaces [1].

In my stimulus talk, I particularly emphasized the second of these topics. Besides, following up on the third topic, in a separate demo session, I demonstrated the integrated prototype (see http://www.phenicx.com) of our recently concluded European PHENICX project.

References

1 Mark S. Melenhorst and Cynthia C. S. Liem. Put the Concert Attendee in the Spotlight. A User-Centered Design and Development approach for Classical Concert Applications. In *Proceedings of the 16th Conference of the International Society for Music Information Retrieval (ISMIR)*, pages 800–806, Málaga, Spain, 2015.

2 Cynthia C. S. Liem, Alessio Bazzica, and Alan Hanjalic. Looking Beyond Sound: Unsupervised Analysis of Musician Videos. In *Proceedings of the 14th International Workshop on Image and Audio Analysis for Multimedia Interactive services (WIAMIS 2013)*, Paris, France, 2013.
3 Cynthia C. S. Liem, Martha A. Larson, and Alan Hanjalic. When Music Makes a Scene – Characterizing Music in Multimedia Contexts via User Scene Descriptions.. *International Journal of Multimedia Information Retrieval*, 2(1): 15–30, March 2013.
4 Cynthia C. S. Liem, Andreas Rauber, Thomas Lidy, Richard Lewis, Christopher Raphael, Joshua D. Reiss, Tim Crawford, and Alan Hanjalic. Music Information Technology and Professional Stakeholder Audiences: Mind the Adoption Gap. In Meinard Müller, Masataka Goto, and Markus Schedl, editors, *Multimodal Music Processing*, Dagstuhl Follow-Ups, volume 3, pages 227–246. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2012. http://www.dagstuhl.de/dagpub/978-3-939897-37-8
5 Cynthia C. S. Liem, Alan Hanjalic, and Craig Stuart Sapp. Expressivity in musical timing in relation to musical structure and interpretation: A cross-performance, audio-based approach. In *Proceedings of the 42nd International AES Conference on Semantic Audio*, pages 255–264, Ilmenau, Germany, July 2011.

3.5 Looking Beneath the Musical Surface

Christopher Raphael (Indiana University – Bloomington, US)

In my talk, I discussed two interrelated music interpretation problems that are both essentially about musical structure. The first deals with the problem of rhythm recognition. While people effortlessly understand complex polyphonic rhythm from audio, in many cases it is nearly impossible for a person or algorithm to correctly perform rhythmic transcription on a sequence of uninflected taps or clicks. The slightest amount of inaccuracy or tempo change can render the sequence ambiguous and "lose" the listener. The reason is that the sequence of times leaves out a great deal of important information that is essential for organizing the music. An interesting formulation of the problem tries to transcribe rhythm given a sequence of onset times now labeled with their pitches, much like what one gets from MIDI stream. The essential human strategy seems to employ a fundamental assumption: when we hear the "same" thing it usually falls in the same position in the measure, beat or other unit in the rhythmic hierarchy. This "same" thing could be a long note, a particular pitch configuration, a short pattern such as a dotted rhythm, etc. This is, of course, the basis for structural analysis by autocorrelation. While this idea is a powerful heuristic it is hard to formulate generally. People are good at recognizing many kinds of variations on the "same" thing though they can involve transformations on many different musical axes, such as leaving out or adding decorative notes, shifting by an interval or chord inversion, contour inversion, reharmonization, etc. The musical understanding seems to require that we identify these variations as a kind of repetition in order to mentally organize the material at an equivalent rhythmic location in the rhythmic structure (e.g. measure). Perhaps the recognition requires that we learn these building blocks or motives in addition to the rhythm we seek to identify. Thus we do not come to each new piece of music we try to understand with a fixed model for rhythm and pitch, but rather recognize by adapting a flexible model to the data and recognizing simultaneously. For instance, a possible recognition strategy could model the music as a sequence of measures with one of several possible types. Standard HMM parameter estimation strategies could be employed to learn the pitch/rhythm tendencies of each of these possible types while parsing the data according to these types. As we seek parsimonious explanations we penalize our models to favor few types of measures. Aside from any specific formulation, the essential idea is to recognize and model the music simultaneously.

An interesting and related problem is that of performing structural and motivic analysis on a simple folk song, carol, anthem, etc, given in score form (rather than performance data). While there could be many motivations for such analysis, one would be to create algorithms that compose simple music that "makes sense" – this might avoid the repetitive music in many computer games, instead providing an inexhaustible supply. As always, there is a close relationship between analysis and synthesis, so it isn't too far fetched to suppose that an analysis engine could form the heart of an algorithmic composition system. Analysis of even the simplest familiar folk song shows a tight reuse of figures and motives at various time scales, while allowing for many possible variations. An algorithmic analysis might try to represent the music in terms of simple building blocks, encoding the music in terms of these basic figures as well as the necessary transformations that lead to the actual notes and rhythm. One could view this approach as analysis by compression, where we seek the minimal number of bits necessary to represent the music. We doubt the power of a strict Lempel-Ziv compression seeking to represent music in terms of literal repetition, since repetition in music is often not exact repetition. Rather one must learn the basic musical objects as well as the transformations applied to them that lead to the musical surface. In formulating the problem scientifically, it may begin with the familiar parsing through a probabilistic context free grammar, though, as before, we do not believe in a generic grammar for music. Rather, as with the previous problem, each piece has its own grammar which must also be learned as part of the recognition strategy. How could one formulate this problem in a way that is both musically plausible and tractable?

3.6 Computational Music Structure Analysis: A Computational Enterprise into Time in Music

Anja Volk (Utrecht University, NL)

In this talk I addressed the relation between computational music structure analysis and the study of time in music. Computational music structure analysis requires the modeling of time processes in music. Structures such as segments, salient patterns, rhythmic-metric structures, and the like are inferred from either symbolic or audio musical content, taking temporal information into account. Often it is crucial for the success of a computational model for solving a specific task within music structure analysis *what* temporal information is taken into account, such as whether large scale or local temporal information is considered. However, even from the perspective of the human information processing in music, we often know only very little about what temporal information we need to consider in a certain context. While music has been argued to be the "art of time," most theories of music have been concerned predominantly with pitch and not with time, such that we are far from understanding the different functions of time and temporal information in music [5].

Time is not only a challenging concept in music analysis, but in many disciplines: what is time, how do we perceive time and how do we successfully employ time in our interactions with the world? While we do not have a sensory organ for perceiving time (as we have, for instance, for colors), studies in cognition demonstrate that the auditory domain is superior to other domains in processing temporal information [4], backing up that it is worthwhile to investigate the different relations between music and time. What is the relation between time as employed in the musical structure and our experience of time when we listen to and make sense of music?

Understanding this relation would help us to employ temporal information in computational music structure analysis in such a way as to find meaningful elements of musical structure, such as temporal building blocks. Current challenges regarding temporal information and music structure analysis discussed during the seminar link to the question of what is the interrelations between different temporal scales in music, such as between temporal microstructure [3] (as studied in expressive timing), small-to-medium-scale temporal phenomena (as studied in the area of rhythm and meter [6]), and medium-to-large scale temporal phenomena (as studied in theories of musical form [2])? For instance, as discussed during the seminar, automatically generating large scale music structures provides an unsolved issue in automatic composition. In the area of rhythm and meter, we know little about the interaction between rhythmic structures and other parts of musical structures, such as melodic patterns. I discussed examples on the role of time for improving music structure analysis for tasks such as repetition-based finding of segments [7], and the discovery of shared patterns in a collection of Dutch folk songs [1]. Hence, computational music structure analysis can help to elucidate the role of time in music for recognizing structure, as well as it will benefit from a better understanding of the human processing of time in music.

References

1 Boot, P., Volk, A., and de Haas, B. (2016). Evaluating the Role of Repeated Patterns in Folk Song Classification and Compression. Submitted for publication.
2 Caplin, W. E., Hepokoski, J., and Webster, J. (2009). Musical Form, Forms and Formenlehre: Three Methodological Reflections, P. Bergé (ed.), Cornell University Press, New York.
3 Clarke, E. (1999). Rhythm and timing in music. In Deutsch, D.: The Psychology of Music, second edition, University of California, San Diego, 473–500.
4 Grahn, J. A., Molly J. Henry, J. H., and McAuley, J. D. (2011). MRI investigation of cross-modal interactions in beat perception: Audition primes vision, but not vice versa. In *NeuroImage*, 54:1231–1243.
5 Kramer, J. D. (1988). The Time of Music. New York: Schirmer Books.
6 London, J. (2004). Hearing in Time, Second Edition. Oxford University Press.
7 Rodriguez-Lopez, M. E. and Volk, A. (2015). Location Constraints for Repetition-Based Segmentation of Melodies. In *Proceedings of the 5th International Conference in Mathematics and Computation in Music (MCM)*, 73–84.

3.7 Defining the Emcee's Flow

Mitchell Ohriner (Shenandoah University – Winchester, US)

In discourse on rap music, the word "flow" takes on several meanings. On the one hand, the flow of a verse is understood as all the musical features of the emcees rapping that verse—rhythms, accents, rhyme, articulation, etc. Flow might also refer to less musical features such as word choice, topic, or reputation. On the other hand, emcees and critics alike attest that individual emcees have a distinctive flow that transcends individual verses or tracks. At the Dagstuhl seminar, using the emcee Black Thought as an example, I presented a method for locating an emcee in a feature space through a large corpus of his verses in comparison to another corpus representative of rap music generally.

References

1 Adams, K. (2009). On the metrical techniques of flow in rap music. Music Theory Online 15(5).
2 Adams, K. (2015). The musical analysis of hip-hop. In Justin Williams (ed.), The Cambridge Companion to Hip-Hop, pp. 118–135. Cambridge University Press.
3 Condit-Schultz, N. (2016). MCFlow: A digital corpus of rap transcriptions. Empirical Musicology Review 11(1).
4 Kautny, O. (2015). Lyrics and flow in rap music. In Justin Williams (ed.), The Cambridge Companion to Hip-Hop, pp. 101–117. Cambridge University Press.
5 Krims, A. (2004). The musical poetics of a 'revolutionary' identity. In Rap Music and the Poetics of Identity. Cambridge University Press.
6 Ohriner, M. (2013). Groove, variety, and disjuncture in the rap of Eminem, André, and Big Boi. Paper present at the Annual Meeting of the Society for Music Theory, November 2, 2013, Charlotte, NC USA.
7 Ohriner, M. (2016). Metric ambiguity and flow in rap music: A corpus-assisted study of Outkast's 'Mainstream' (1996). Empirical Musicology Review 11(1).

3.8 Richard Wagner's Concept of 'Poetico-Musical Period' as a Hypothesis for Computer-Based Harmonic Analysis

Rainer Kleinertz (Universität des Saarlandes, DE)

Joint work of Meinard Müller, Christof Weiß, Rainer Kleinertz

In the third part of his large theoretical work *Oper und Drama* (Leipzig 1852), Wagner developed the idea of a 'poetico-musical period.' Based on the *drama* (i.e. text and action), he tries to motivate modulations: The *musician* (composer) would receive an incitement to step outside the once selected key only when an opposite emotion occurs (e.g., "Die Liebe bringt Lust und Leid."). When this new, opposite emotion returns to the original emotion (e.g., "Doch in ihr Weh auch webt sie Wonnen."), then harmony would return in the original key. In Wagner's eyes the most perfect artwork would be that, in which many such 'poetico-musical periods' – as he calls them – present themselves "in utmost fulness."

These pages on his *Drama of the Future* were applied to Wagner's musical dramas by Alfred Lorenz in his highly influential study *Das Geheimnis der Form bei Richard Wagner* (4 vols., Berlin 1924–1933). Among others, Lorenz analyzed the entire Ring as a series of such periods. In the 1960s, Carl Dahlhaus rejected Lorenz' analyses as being completely erroneous and against Wagner's musical ideas. In 2002 Werner Breig postulated that the concept should be ignored, as Wagner – when he coined it – had not yet composed a single note of the Ring, and even the texts of *Rheingold*, *Die Walküre*, and *Siegfried* did not yet exist. My own hypothesis – as published in [2] – is that Wagner had indeed something in mind which he realized at least partly in his subsequent Ring composition: The concept of the 'poetico-musical period' serves to describe more or less 'closed' parts of the Ring in which a strong emotional change motivates modulations leaving the original key and returning to it. As a paradigmatical example for such a ‘poetico-musical period’ may serve Sieglinde’s narration in the first act of *Die Walküre* with its two interior modulations out of and back to the framing tonality of E minor. Consequently, the ‘poetico-musical period’ should not be regarded as a mere ‘way’ of music between a certain tonality and its return, but as a harmonic construct around a central modulation.

This musical-philological assumption of what Wagner may have had in mind when he wrote *Oper und Drama* may serve as a meaningful hypothesis for computer-based harmonical analysis. In a current research project of the Deutsche Forschungsgemeinschaft (DFG) – a cooperation of Meinard Müller’s group in Erlangen and Rainer Kleinertz’ group in Saarbrücken – harmonic analysis of the entire Ring based on audio data may allow a verification or falsification of this hypothesis. This approach may become a paradigm for a cooperation between historical musicology and computer science where the fundamentally different methods of both disciplines are applied in favour of new *objectified* results. Hermeneutical-musicological understanding and computer-based proceedings would allow new insights in complex musical works.

References

1 Verena Konz, Meinard Müller and Rainer Kleinertz (2013). A Cross-Version Chord Labelling Approach for Exploring Harmonic Structures – A Case Study on Beethoven’s Appassionata, Journal of New Music Research, DOI: 10.1080/09298215.2012.750369
2 Rainer Kleinertz (2014). Richard Wagners Begriff der ‘dichterisch-musikalischen Periode’, Die Musikforschung 67, pp. 26–47.

3.9 Large-Scale Structures in Computer-Generated Music

Mary Farbood (New York University, US)

The generation of musical structure is closely related to (and in most cases dependent on) the analysis of structure. Short-term musical structure has been effectively modeled in many algorithmic composition systems ranging from those that compose music off-line such as David Cope’s EMI to improvisatory systems such as François Pachet’s Continuator. These systems often use Markov models or some type of generative grammar to create longer sections of music from shorter segments or individual notes. The success of these systems in composing convincingly human-sounding music is dependent on how well lower-level generation and assembly of smaller segments make stylistic sense. Implementing a system that produces convincing computer-generated music formalizes (at least to some extent) how listeners intuitively perceive style. However, computer-generated music that is coherent and interesting on a large-scale structural level is very difficult to achieve. No system currently exists that produces aesthetically compelling music (at least from a structural perspective) over long time spans. At Dagstuhl, we discussed this problem in the context of cognitive constraints and aesthetic considerations.

3.10 A Composer's Perspective on MIR

Carmine-Emanuele Cella (École normale supérieure, F)

A sound transformation is, in a general sense, any process that changes or alters a sound in a significant way. Transformations are closely related to representations: each action is, indeed, performed on a specific representation level. For example, a time stretch performed in the time domain by means of granular synthesis gives inferior results (perceptually) to the same transformation performed in the frequency domain, where one has access to phase information. In the same way, a pitch shift operated in the frequency domain gives inferior results to the same operation performed using a spectral envelope representation, where one has access to dominant regions in the signal.

In the two cases discussed above, we passed from a low-level representation (waveform) to a middle-level representation (spectral envelope). We could, ideally, iterate this process by increasing the level of abstraction in a representation, thus giving access to specific properties of sound that are perceptually relevant; by means of a powerful representation it could therefore be possible to access a semantic level for transformations.

An example will clarify the ideas outlined: suppose we want to elongate the minor chords present in a sound by a certain factor, but only if they are located in non-transient regions. At the same time, we want to pitch shift them by some semitones, but only if they are played by a piano. Obviously, this kind of textual description is very easy to understand by humans, but extremely difficult to code in an algorithm. We envision, therefore, the possibility in the future to have such kind of semantic transformations.

3.11 Evolution and Salience

Geraint A. Wiggins (Goldsmiths, University of London, GB)

My work is currently focused on a cognitive architecture that is intended to explain the structuring of sequential and semantic information in the mind/brain. Because it is fundamentally sequential, it is directly applicable to music and language. One aspect of the approach is that it ought to explain what we mean by "salience" in our musicological discussions, in line with the paradigmatic analytical approach of Ruwet, and the related psychological theory of cue abstraction, due to Deliège [1]. There are several ways to look at music, and they can tell us different things. One way is via evolution. When I consider music this way, in the context of the information-processing view of mind, I am led to the notion of salience, which is something we don't often discuss in MIR.

References

1 G. A. Wiggins. Cue abstraction, paradigmatic analysis and information dynamics: Towards music analysis by cognitive model. Musicae Scientiae, Special Issue: Understanding musical structure and form: papers in honour of Irène Deliège, pages 307–322, 2010.

3.12 Beat Tracking with Music Structure

Roger B. Dannenberg (Carnegie Mellon University, US)

I want to encourage more thinking about music structure as the "missing link" in music understanding. The MIR community has made great progress in using larger datasets, machine learning, and careful search and optimization to improve the performance of many music understanding tasks, including beat tracking. However, our algorithms still make obvious and silly mistakes, at least when viewed from a human perspective. My sense is that most current algorithms for music understanding tasks are very selective about the information they use. The information is highly effective for most problems, which is why systems work at all, but when much of the available information is ignored, algorithms can never be robust.

I believe that music is intentionally confusing and created in a way that demands attention to many different aspects of rhythm, melody, and harmony. It is not clear why music works this way, but it seems natural that music would exercise human intelligence, and the brain does seem to enjoy making sense out of things, especially when they are not completely obvious. If this is the way music works, then we must think about integrating many sources of information in order to get our machines to understand music. Music understanding is largely a problem of finding patterns and structure in music. For example, if we can identify the structure of conventional music in terms of phrases and measures, most of the work of beat tracking is done.

With this premise in mind, let us consider how we might use non-beat information to help with the beat-tracking problem. This approach is mainly a review of an earlier ISMIR paper [1], considered here in the context of this seminar on music structure. Nearly all beat trackers optimize two basic things:

- Beats and "beat features" are highly correlated. In other words, something in the signal (amplitude change, increase in high frequencies, spectral difference) indicates where beats are likely to occur.
- Tempo is fairly steady. In other words, the spacing between nearby beats is about the same.

These constraints are expressed differently in different beat tracking algorithms, but seem to be the core principles.

I propose a third principle based on music structure that can provide additional information: Where repetitions occur in music, the beats in the two repetitions should correspond. For example, suppose a beat tracker labels a verse correctly at the beginning of a song and the verse repeats later. Rather than relabeling the same music (and possibly getting it wrong), this constraint tells us that the beats in the two sections should correspond; thus, we have essentially labeled the beats just by identifying the structural repetition. Alternatively, we can process these two repetitions as one. Then, we have twice as much information to help us find the beats. Of course, the information may be conflicting, but generally it should be easier to find one solution than two.

An implementation of this approach uses a self-similarity matrix based on chroma features to find repetition in music. The result of this step is a set of continuous mappings from one interval of time to another. There is one mapping for each repetition that is discovered. Then, each time a beat is proposed at time t, we simply map t to all the places this beat should be repeated and propose there must be beats at those locations as well. In practice, chroma-based alignments do not have high time resolution, so the mappings are not very useful for placing single beats. However, we can modify the constraint to say that tempo is consistent in repeated sections. Thus, when we place several measures of beats in one location, we can use the mappings to assert the tempo in other locations. The tempo may not be identical due to slight tempo changes during the performance, but the derivatives of the mappings can be used to estimate the tempo change.

In practice, these constraints have been imposed through a gradient descent algorithm. Essentially, a few beats are placed according to initial guesses based on the autocorrelation of beat features. These guesses are replicated according to repetitions in the music. Then, beat locations are adjusted to optimize the combination of all three constraints: beats correspond to beat features, tempo is steady, and tempo is consistent across repetitions. The algorithm continues by alternately proposing a few more beats (based on the steady-tempo principle) and then optimizing using gradient descent, until the entire piece is covered by beats.

This is only one approach, presented to motivate thinking about ways we can use music structure in music processing tasks. One of the problems with this approach in general is that while structure can offer helpful information, the information can also be wrong and misleading. Building more holistic algorithms does not guarantee improvement over simpler approaches that benefit from being more tractable and more amenable to training with large datasets. Another challenge is to jointly estimate structure along with information such as tempo, beats, rhythm, and harmony.

References

1 Roger B. Dannenberg. *Toward Automated Holistic Beat Tracking, Music Analysis, and Understanding* in Proceedings of the 6th International Conference on Music Information Retrieval Proceedings (ISMIR), London, pages 366–373, 2005.

3.13 Can We Reach a Consensus on the Minimum Amount of Originality to Regard a Piece of Music as Original?

Masataka Goto (AIST – Tsukuba, JP)

In the age of digital music, future musicians may find it more difficult to be truly original in the face of ever-expanding archives of all past music. The amount of digital musical pieces that can be accessed by people has been increasing and will continue to do so in the future. Since the amount of similar musical pieces is monotonically increasing, musicians will be more concerned that their pieces might invite unwarranted suspicion of plagiarism. All kinds of musical pieces are influenced by existing pieces, and it is difficult to avoid the unconscious creation of music partly similar in some way to prior music. The monotonic increase in musical pieces thus means that there is a growing risk that one's piece will be denounced as being similar to someone else's.

To address this issue, we started a research project called OngaCREST [1] to build an information environment in which people can know the answers to the questions "What is similar here?" and "How often does this occur?" Although human ability to detect musical similarity and commonness (typicality) [2] is limited for a large-scale music collection, future advanced technologies would enable people to compute musical similarity between any pairs of musical pieces and musical commonness of a musical piece to a set of pieces.

Once such technologies could be available to compute musical similarity and commonness in detail, people could naturally understand that any musical piece has similar pieces with regard to some aspects. The concept of originality would then be discussed in a more quantitative way and might be revised. If some (or most) aspects are always similar, how can we measure the amount of originality? To be regarded as an original musical piece, how many different aspects or elements should it have? Can we, as a global society, reach a consensus on the minimum amount of originality to regard a piece of music as original?

References

1 Masataka Goto. *Frontiers of Music Information Research Based on Signal Processing*, Proceedings of the 12th IEEE International Conference on Signal Processing (ICSP), pages 7–14, October 2014.

2 Tomoyasu Nakano, Kazuyoshi Yoshii, and Masataka Goto. *Musical Similarity and Commonness Estimation Based on Probabilistic Generative Models*, Proceedings of the IEEE International Symposium on Multimedia (ISM), pages 197–204, December 2015.

3.14 Let's Untie Our Hands! Use All the Data You Have and Stop Making Life Difficult

Mark Sandler (Queen Mary University of London, GB)

URL http://www.semanticaudio.ac.uk/

At the Dagstuhl seminar, I said a few words on music structure analysis in the context of a large project called "Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption" (FAST-IMPACt) I am coordinating. Thinking about musical structure, one needs to ask, who is this for? Is it for the professional in the studio (a strong focus of FAST-IMPACt) or the consumer, or even for some intermediary needing to make money somehow from the content – hopefully on behalf of the artists and creators? The needs of different categories of user are very different. For the producer it is probably connected with navigation around a particular piece of music in an ongoing project. For the consumer/listener this could be true, but there is potentially the added need to navigate within collections. (As I write I realise that the latter is also true for professionals, though the collections are different!) Taking a step back, we can say that for many musics, what we need to do to help these participants is analyse the audio signal and extract meaningful, and above all, useful information from the audio. We need to do this to the best of the capabilities of the available technologies and algorithms. Everyone would find this hard to dispute, I think. Yet, why do we all, to my knowledge with zero (or close to) exception, make use of anything but a monophonic down mix? I would therefore propose that we start to investigate ways that use the maximum amount of data and information available to us, and to stop making our investigations overly and unnecessarily difficult. I would start with stereo signals – which I see some researchers describe as 'legacy'!

3.15 Towards an Information-Theoretic Framework for Music Structure

Frédéric Bimbot (CNRS/IRISA, Rennes, FR)

Music is a communication signal and the estimation of music structure is essentially an information-theoretic problem. The structure S of a music content M can be understood as the "proper type" and "right quantity" of latent side information which provides an economical explanation of M by minimizing $Q(M, S)$, i.e., the quantity of information needed to jointly describe M and S. Two philosophies can support the definition of Q (see for instance [4]):

- Shannon's Information (SI), also called lossy source-coding scheme, which relates information Q to the distortion of M with respect to a prototypical structure itself derived from a probabilistic model of all possible structures, and
- Kolmogorov's Complexity (KC), sometimes referred to as algorithmic compressibility, which considers M as the output of a short, standalone program (within a class of valid structure generating programs), whose size is related to Q.

Shannon's approach is fundamentally a knowledge-based (inter-opus) approach, where statistically typical forms provide templates that guide the recognition of music content organization (stylistic structure). Kolmogorov's framework is rather based on a data-driven (intra-opus) viewpoint and focuses on internal redundancy as a primary criterion for grouping musical material into consistent structural patterns ("semiotic" structure [1, 2]). Both conceptions of information are meaningful, but understanding and exploiting their interaction remains a fundamental scientific bottleneck – in MIR, in Computational Musicology, and also in many other scientific domains. The duality between SI and KC in music is for instance illustrated by Schenker's [10] versus Narmour's [7, 8] conceptions of music structure, and KC approaches are becoming increasingly popular in MIR (see for instance [5]).

However, current approaches in Music Structure Analysis [6] fail in explicitly accounting for both aspects simultaneously, even though they are presumably present with a different balance across musical genres (this could be one of the causes of ambiguities in human perception of structure [11]). Note that, even though, neither SI nor KC can actually be calculated exactly, they can be estimated using models, i.e. family of distributions for SI such as Hidden Markov Models (see for instance [9]) and classes of programs for KC (as prefigured by the System & Contrast Model [3]). Approaching the diverse views of music structure within the common framework of Information Theory appears as a relevant move towards a better understanding of what music structure is, but also as a key for a more efficient use of music structure in computational contexts. Though music is not a uniquely decodable code, it can be assumed that the number of reasonable structural hypothesis is sufficiently limited so as to be tackled in a relatively unified framework, encompassing the two main facets of data compression (SI and KC). By understanding the interaction between the "two sides of a same coin," music could become a case study that would help bridging a fundamental gap in Information Sciences.

References

1 Bimbot, F., Deruty, E., Sargent, G., Vincent E. (2012). Semiotic Structure Labeling of Music Pieces: Concepts, Methods and Annotation Conventions. Proc. International Society on Music Information Retrieval Conference (ISMIR), pages 235–240, Porto.

2 Bimbot, F., Sargent, G., Deruty, E., Guichaoua, C., Vincent, E. (2014). Semiotic Description of Music Structure: An Introduction to the Quaero/Metiss Structural Annotations. Proc. 53rd AES Conference on Semantic Audio, 12 pages, London.
3 Bimbot, F., Deruty, E., Sargent, G., Vincent, E. (2016). System & Contrast: A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces. Music Perception. 41 pages, to appear.
4 Grünwald, P., Vitanyi, P. (2004). Shannon Information and Kolmogorov Complexity. arXiv preprint `cs/0410002. 2004`, updated 2010.
5 Meredith, D. (2012). Music Analysis and Kolmogorov Complexity. Proc. XIX CIM.
6 Müller, M. (2015). Music structure analysis. In Fundamentals of Music Processing, chapter 4, pages 167–236, Springer Verlag.
7 Narmour, E. (1977). Beyond Schenkerism. University of Chicago Press.
8 Narmour, E. (2000). Music expectation by cognitive rule-mapping. Music Perception, XVII/3, pages 329–398.
9 Pauwels, J., Peeters, G. (2013). Segmenting music through the joint estimation of keys, chords and structural boundaries. Proc. 21st ACM International Conference on Multimedia, New York.
10 Schenker, H. (1935). Der freier Satz, Universal, Vienna.
11 Smith, J. (2014). Explaining Listener Differences in the Perception of Musical Structure. PhD thesis, Queen Mary University of London, UK.

3.16 MorpheuS: Constraining Structure in Music Generation

Dorien Herremans, Elaine Chew (Queen Mary University of London, GB)

A major problem with much of the automatically generated music is that it lacks a structure and long-term coherence. We have defined the music generation problem as a combinatorial optimization problem [2, 5]. The advantage of this approach is that it gives us the freedom to impose both hard and soft constraints. These constraints can be used to define different types of structure.

One example of a structure that can be imposed by hard constraints is based on repeated and transposed patterns. The cosiatec pattern detection algorithm [3] was used to find maximum translatable patterns. These patterns were then used to constrain the output of a music generation algorithm called MorpheuS (http://dorienherremans.com/software).

A second form of structure, which is soft constrained, is a tension profile. This type of tension could be relevant to, for instance, automatic generation of game or video music. We have developed a model [4] that captures aspects of tonal tension based on the spiral array [1], a three dimensional model for tonality. Our approach first segments a musical excerpt into equal length subdivisions and maps the notes to clouds of points in the spiral array. Using vector-based methods, four aspects of tonal tension are identified from these clouds. First, the cloud diameter measures the dispersion of clusters of notes in tonal space. Second, the cloud momentum measures the movement of pitch sets in the spiral array. Third, the tensile strain measures the distance between the local and global tonal context. Finally, the cosine similarity measures the directional change for movements in tonal space.

The results of generating polyphonic piano music with constrained patterns and fit to a tension profile are very promising and sound musically interesting. The reader is invited to listen to full pieces generated by the algorithm at http://dorienherremans.com/MorpheuS.

References

1. Chew, Elaine (2014). *The Spiral Array*. Mathematical and Computational Modeling of Tonality. Springer, pages 41–60.
2. Herremans, Dorien, and Sörensen, Kenneth (2013). *Composing fifth species counterpoint music with a variable neighborhood search algorithm*. Expert Systems with Applications 40(16):6427–6437.
3. Meredith, David (2013). *COSIATEC and SIATECCompress: Pattern discovery by geometric compression*. in Music Information Retrieval Evaluation Exchange (Competition on "Discovery of Repeated Themes & Sections") of the International Society for Music Information Retrieval Conference.
4. Herremans, D., and Chew, E. (2016). *Tension ribbons: Quantifying and visualising tonal tension*. Second International Conference on Technologies for Music Notation and Representation (TENOR). Cambridge, UK.
5. Herremans, D., Weisser, S., Sörensen, K., and Conklin, D. (2015). *Generating structured music for bagana using quality metrics based on Markov models*. Expert Systems with Applications 42(21):7424–7435.

3.17 Using Prior Expectations to Improve Structural Analysis: A Cautionary Tale

Jordan Smith (AIST – Tsukuba, JP)

Joint work of Masataka Goto, Jordan Smith

Annotations of musical structure tend to have strong regularities: average segment size is roughly 20 seconds, the number of segments per annotation is roughly 12, and segments usually have a uniform size, meaning that the average ratio of segment length to the median segment length is very close to 1. These regularities are consistent even between different collections of annotations, but are not often used by algorithms to refine estimates. By treating these regularities as prior expectations, we can use a committee-based approach to structural analysis: first, make several estimates using a variety of algorithms; second, choose the estimate that is likeliest given the prior distributions. Although the method may seem like 'cheating' (even when appropriate leave-one-out training regimes are followed), the approach is guaranteed to give at least a modest gain in f-measure.

Except, we tried it, and it didn't work. Why not? We are still trying to decide.

3.18 Music Segmentation: of what, for what, for who

Xavier Serra (UPF – Barcelona, ES)

Within the field of Computational Musicology, we study music through its digital traces, or digital artifacts, that we can process computationally. Ideally we want to start from data that is as much structured as possible and the goal is to extract musical knowledge from it. However most current computational research is still focused on trying to increase the structuring of the existing data by computational means. Music segmentation is a key structuring element and thus an important task is to do this as automatically as possible.

Most, if not all, music data processing problems, and music segmentation is no exception, should be approached by taking into consideration the following issues: What signal and music are we processing? What application are we aiming at? Who is the user and context being targeted?

Each type of music signal (including audio, symbolic scores, lyrics, etc.) requires different segmentation methodologies and implies different segmentation concepts. Each musical facet (including timbre, melody, rhythm, harmony, etc.) also requires a different music segmentation strategy and can be used for different tasks.

The targeted application of a music segmentation process is also critical. It is very different wanting to perform music analysis for music understanding or wanting to solve some engineering task-driven problem. It is nice when a task-driven problem can be based on a musically grounded concept, but it is not always possible, nor even adequate.

Personal subjectivity, cultural bias, and other contextual issues greatly affect the concept of music segmentation. The analysis approach has to take that into account and assume that the results obtained should be different depending on the context being targeted.

In general, the concept of music segmentation means many things, even within the MIR community. Maybe the most common meaning relates to music structure, which is a musically grounded concept. But strictly speaking, practically all music processing tasks have an implicit or explicit segmentation, and this segmentation has a big impact on the results obtained. We process music signals by first segmenting them into discrete events, such as audio samples, audio frames, symbolic notes, phrases, songs, and so on. We use a multiplicity of segments with more or less agreed definitions and standard approaches to be computed. Clearly any particular piece of music has many possible multilevel segmentations that might be of use for different applications and different contexts.

As a summary I want to emphasize that we cannot talk about music segmentation without taking into account

- the type of signals we start from,
- the targeted application, and
- the particular context in which the signal and application is part of.

In the project CompMusic (see http://compmusic.upf.edu), we have worried about these three issues. Let me go through what has been our approach.

With respect to the issue “of what,” we have focused on the musical repertories of five music cultures: Hindustani, Carnatic, Turkish-makam, Beijing Opera, and Arab-Andalusian. The data we have been processing has been mainly audio recordings, scores, and editorial metadata. We also have extracted audio features from the audio recordings that are then used as inputs to other processing tasks. Segmenting the audio and the scores in a unified way has been an important task in some of the music repertories.

With respect to the issue "for what," our main goal has been to develop tools and systems with which to explore the selected musical repertories. Dunya (see http://dunya.compmusic.upf.edu) is one of the prototypes we have developed in order to evaluate the developed technologies, a prototype that can be used to explore the music collections we have compiled and with it you can listen to music pieces while visualizing information that can help the understanding and enjoyment of the music.

With respect to the issue "for who," we have aimed at developing tools that can be of interest to the music lovers of each of the music traditions we have been studying. Thus we target people with some knowledge of the music they are exploring and listening to.

3.19 Flexible Frameworks for the Analysis of Rhythm and Meter in Music

Andre Holzapfel (OFAI-Wien, AT)

In a recent study [1] on metered Turkish makam music performance, we illustrated differences in the ways notes are distributed in compositions from different cultures. Based on these findings, we are now able to track meters that go beyond simple meters in 4/4 and 3/4. However, our evaluation measures are tailored towards simple/symmetric meters, our evaluation data is limited in style, annotations are mostly from one annotator, and in most cases only the beat is annotated, while other metrical layers are ignored. Recent developments both in Bayesian Networks and Deep Neural Networks push the state of the art in meter tracking to a new level. However, how far can we go given the limited amounts of annotated data, and possibly more importantly, the limited amount of understanding of musics that we engineers have about the diverse structures we aim to subject to a correct analysis? We developed a Bayesian framework [2] for meter tracking in music that is able to track meter, given a small amount of annotated representative samples. A Bayesian framework allows to adapt the model to new features, and to different types of tempo and meter properties. In a discussion of important steps to take in future, I would like to emphasize that including a complete set of observables into account is highly timely; Learning meter in Indian music without looking at performers seems odd. Furthermore, music performances are shaped by humans who move and breathe together, and the aspects in which their various biosignals correlate remain widely unknown. In short, we need methodologies for the sophisticated observation of performance events. But still, they will not reveal the meanings and mental representations that these structures are evoke in various contexts. I believe that in this aspect interdisciplinary collaborations between music psychology, engineering, and ethnomusicology can indicate promising directions that go beyond observationalism towards a more complete understanding of music, driven by sophisticated computational analysis.

References

1 Holzapfel, A. (2015). Relation between surface rhythm and rhythmic modes in Turkish makam music. Journal for New Music Research 44(1):25–38.
2 Krebs, F., Holzapfel, A., Cemgil, A. T., and Widmer, G (2015). Inferring metrical structure in music using particle filters. IEEE/ACM Transactions on Audio, Speech and Language Processing, 23(5):817–827.

3.20 Music Structure: Scale, Homegeneity/Repetition, Musical Knowledge

Geoffroy Peeters (UMR STMS – Paris, FR)

Joint work of Florian Kaiser, Johan Pauwels, Geoffroy Peeters

In my talk at the Dagstuhl seminar, I discussed three aspects of automatic music structure estimation.

The first aspect relates to the temporal scale considered when trying to estimate the structure (i.e., the duration of the segments). This scale is usually a priori unknown. To solve this issue, we proposed in [1] to use a multi-scale approach in which a set of checkerboard-kernels of increasing size is used to segment a given self-similarity matrix.

The second aspect relates to the underlying process that creates the structure. Currently two major assumptions are used: homogeneity/novelty and repetition [3] leading to the so-called "state" and "sequence" approaches [5]. Also, this underlying process is typically a priori unknown. To solve this issue, we proposed in [2] a joint estimation based on the two assumptions leading to a large increase in the estimation results.

Finally, I discussed how musical structure can be estimated exploiting musical knowledge. As an example, I reviewed our work [4] on the joint estimation of chord, key, and structure, were the structure arises from the variation of chord perplexity at the end of each segment.

References

1 F. Kaiser and G. Peeters. Multiple hypotheses at multiple scales for audio novelty computation within music. In Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vancouver, British Columbia, Canada, pages 231–235, 2013.

2 F. Kaiser and G. Peeters. A simple fusion method of state and sequence segmentation for music structure discovery. In Proc. of the International Society for Music Information Retrieval Conference (ISMIR), Curitiba, PR, Brazil, pages 257–262, 2013.

3 J. Paulus, M. Müller, and A. Klapuri. Audio-based Music Structure Analysis. In Proc. of the International Conference on Music Information Retrieval (ISMIR), Utrecht, The Netherlands, pages 625–636, 2010.

4 J. Pauwels, F. Kaiser, and G. Peeters. Combining harmony-based and novelty-based approaches for structural segmentation. In Proc. of the International Society for Music Information Retrieval (ISMIR), Curitiba, PR, Brazil, pages 601–606, 2013.

5 G. Peeters. Deriving Musical Structures from Signal Analysis for Music Audio Summary Generation: Sequence and State Approach. In Proc. Computer Music Modeling and Retrieval (CMMR), Lecture Notes in Computer Scienc 2771, Springer, pages 142–165, 2004.

4 Further Topics and Open Problems

4.1 Musical Structure Analysis for Jazz Recordings

Stefan Balke, Meinard Müller (Universität Erlangen-Nürnberg, DE)

Analyzing jazz recordings by famous artists is the basis for many tasks in the field of jazz education and musicology. Although jazz music mainly consists of improvised parts, it follows common structures and conventions which allow musicians to play and interact with each other. For example, at jam sessions and in traditional jazz recordings, musicians introduce a song by playing its main melody based on a characteristic harmonic progression. This part is also called the *head-in*. Afterwards, this progression is repeated while the melody is replaced by improvised solos by the various musicians. After all solos have been played, the song is concluded with another rendition of the main melody, a part also referred to as *head-out*. Based on this musical knowledge, we investigated automated methods for detecting (approximate) repetitions of the harmonic progression, certain melodic elements, and transitions between soloists as cues to derive a coarse structure of the jazz recording. The discussions at the Dagstuhl seminar showed that the integration of specific domain knowledge is essential for dealing with the possible musical and acoustic variations one encounters in jazz music.

References

1 Klaus Frieler, Wolf-Georg Zaddach, Jakob Abeßer, and Martin Pfleiderer. Introducing the Jazzomat Project and the melospy Library. In *Third International Workshop on Folk Music Analysis*, 2013.
2 The Jazzomat Research Project. http://jazzomat.hfm-weimar.de.

4.2 On the Role of Long-Term Structure for the Detection of Short-Term Music Events

Juan Pablo Bello (New York University, US)

In MIR it is often assumed that there is a universal "ground truth" to music events such as beats, downbeats, chords, and melodic lines. This conveniently ignores the fact that music analysis is an interpretative task, with multiple outcomes possible. But, if there is no single answer, what should we expect from computational approaches?

One possible objective is to produce at least one valid answer, one that could have plausibly been produced by a human. I would argue that plausibility is partly a function of the long-term structural coherence of the system's output, an aspect that is largely ignored during the design, training and evaluation of current approaches. As a result, music event detection is typically performed as a series of short-term, (semi-)independent tasks, with outputs that are often incoherent and thus implausible.

Take for example chord estimation, where methods are trained on maximum likelihood objectives derived from windows of information rarely spanning more than a single chord; dynamic models, whenever used, are almost certain to have short memories; and evaluation is based on an aggregation of the accuracy of short-term detections. Earlier work tried to leverage long-term repetitions to enhance the robustness of feature representations with promising results [1, 2], but those strategies have not been widely adopted, having next to no impact on the feature extraction, model training and evaluation methodologies currently in use.

During the Dagstuhl seminar we have discussed the multiple ways in which the long-term, hierarchical structure of musical pieces can be used to improve the validity, and thus usability, of computational music analyses. However, some of these issues were only discussed briefly and tentatively, and much remains open for future discussion and development within the community.

References

1 Matthias Mauch, Katy Noland, and Simon Dixon. *Using Musical Structure to Enhance Automatic Chord Transcription*. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Kobe, Japan, pages 231–236, 2009.
2 Taemin Cho and Juan P. Bello. *A Feature Smoothing Method for Chord Recognition using Recurrence Plots*. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Miami, USA, pages 651–656, 2011.

4.3 Mid-level Representations for Rhythmic Patterns

Christian Dittmar, Meinard Müller (Universität Erlangen-Nürnberg, DE)

For music retrieval and similarity search, one important step is to convert the music data into suitable mid-level features. Ideally, these representations should capture relevant characteristics of the music while being invariant to aspects irrelevant for the given task. For example, rhythmic patterns that are played in different tempi may be perceived as similar by human listeners, while being numerically quite different. In this context, one requires mid-level representations that capture rhythmic characteristics while being invariant to tempo changes. During the Dagstuhl seminar we revisited different mid-level features that have been proposed in earlier works to capture rhythmic information. An established technique for analyzing rhythmic patterns is based on computing a local version of the autocorrelation function (ACF) of some onset-related function [1]. Together with Andre Holzapfel, we discussed open issues related to applying the scale transform [3] to rhythmic patterns for improving tempo invariance. In a follow-up discussion with Brian McFee, we highlighted the relation between the scale transform and the Log-lag ACF [2]. Together with Frank Kurth, we investigated the suitability of shift-ACF [4] for characterizing rhythmic structures with multiple repetitions.

References

1 Peter Grosche, Meinard Müller, and Frank Kurth. Cyclic tempogram – a mid-level tempo representation for music signals. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 5522–5525, Dallas, Texas, USA, 2010.
2 Matthias Gruhne and Christian Dittmar. Improving rhythmic pattern features based on logarithmic preprocessing. In *Proceedings of the Audio Engineering Society (AES) Convention*, Munich, Germany, 2009.

3 Andre Holzapfel and Yannis Stylianou. Scale transform in rhythmic similarity of music. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(1): 176–185, 2011.
4 Frank Kurth. The Shift-ACF: Detecting multiply repeated signal components. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–4, New Paltz, NY, USA, 2013.

4.4 Representation of Musical Structure for a Computationally Feasible Integration with Audio-Based Methods

Sebastian Ewert (Queen Mary University of London, GB)

In terms of terminology, "musical structure" has been used in several, different contexts. In one interpretation, musical structure is essentially equivalent to musical form, which can be considered as a genre or rather style specific definition of the expectation of how a piece is composed on a rather global level. Another interpretation of structure is closer to the corresponding mathematical notion, where structure yields properties and regularity.

Both interpretations lead to various interesting questions. In the context of the first interpretation, a popular task is to determine the temporal boundaries of elements used to describe the form, such as the chorus in pop music or the recapitulation in a sonata form. One group of methods focuses on the detection of boundaries by detecting novelty or sudden changes in terms of a feature representation. This is essentially equivalent to a local, focused expectation violation of some sort. In this context, we discussed what this means for various music and composition styles, and how this could be expressed computationally using audio recordings as input.

This is directly connected to a question we raised in the context of the second interpretation of structure. Here, structure can refer to various regularities or expectations about the harmony, the rhythm, the melody or any other musical concept. In this context, music signal processing as a field has been criticized for not making enough use of these properties to obtain better results in specific tasks. While this criticism is valid it often leads to simplifying conclusions about the underlying reasons for why structure is neglected. In particular, a major obstacle is that the detection of musical low-level events is still an unsolved problem (e.g. note transcription). Therefore, good signal processing methods typically avoid making hard decisions (e.g. "This is a C-major chord") but preserve uncertainty in a model as long as possible. This, however, leads to an exponential explosion of the underlying state space for longer time ranges and, therefore, we simply often cannot represent or integrate complex expectation models that require long time ranges – at least not using classical, symbolic Bayesian modelling techniques.

Recently, neural networks, in contrast to attempts modelling expectations explicitly, express expectations implicitly and thus can be used to build complex language models (i.e. expectation models) for polyphonic music, even down to a note level. This led to measurable but small improvements in tasks such as music transcription, and thus can be considered as a first step. A more recently developed mathematical tool are uncertainty functions, which avoid an explosion of the state space similar to neural network language models but at the same time enable the integration of explicit knowledge (to some degree). In this context, wed discussed approaches and best practices to representing musical structure for specific (!) cases, where the Bayesian network philosophy fails – in particular with respect to usable, practical ways for integrating such representations into audio signal processing methods while preserving computational feasibility.

4.5 Robust Features for Representing Structured Signal Components

Frank Kurth (Fraunhofer FKIE – Wachtberg, DE)

In the last years we have developed several features for robustly representing repeating signal components [1]. A main focus in this was the robust detection of such signal components in real-world audio recordings. Applications included bioacoustical monitoring [2] (e.g., detection of repeating bird calls or sounds of marine mammals) and speech detection [3]. In the latter, the harmonic structure of voiced signal parts constitute the repeating components. One of my interests during the Dagstuhl seminar was to discuss possible applications of such features in music structure analysis.

Generally speaking, repetitions can be seen as building blocks for more complex structure elements of audio signals, which is particularly obvious for music. Thus, another interesting thing discussed at the seminar was that of possible generalizations of the repetition-based features proposed in [2] to represent such complex structures. As a third possible application, the usability of such features to the extraction and separation of mixtures of repeating components (e.g, for multipitch extraction [3] or the detection of overlapping rhythmic components) was discussed.

References

1 Frank Kurth. *The shift-ACF: Detecting multiply repeated signal components.* In Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 2013.
2 Paul M. Baggenstoss and Frank Kurth. *Comparing Shift-ACF with Cepstrum for Detection of Burst Pulses in Impulsive Noise*, Journal of the Acoustical Society of America 136(4):1574–1582, 2014.
3 Alessia Cornaggia-Urrigshardt and Frank Kurth. *Using enhanced F0-trajectories for Multiple Speaker Detection in Audio Monitoring Scenarios.* In Proc. of the European Signal Processing Conference (EUSIPCO), Nice, France, pages 1093–1097, 2015.

4.6 Reversing the Music Structure Analysis Problem

Meinard Müller (Universität Erlangen-Nürnberg, DE)

The general goal of music structure analysis is to divide a given music representation into temporal segments that correspond to musical parts and to group these segments into musically meaningful categories [1, 2]. In general, there are many different criteria for segmenting and structuring music. For example, a musical structure may be related to recurring patterns such as repeating sections. Or a certain musical sections may be characterized by some homogeneity property such as a consistent timbre, the presence of a specific instrument, or the usage of certain harmonies. Furthermore, segment boundaries may go along with sudden changes in musical properties such as tempo, dynamics, or the musical key [1]. When recognizing and deriving structural information, humans seem to combine different segmentation cues in an adaptive and subjective fashion [3]. The listener-dependent and context-sensitive relevance of different segmentation principles make structure analysis an extremely challenging task when approached with computer-based systems. During the Dagstuhl seminar, we discussed a task that may be regarded as a kind of reversal of the structure analysis problem: Given a structure annotation made by a human listener, find out possible segmentation cues that support the annotation. A similar task was suggested by Smith and Chew [4], where a given structure annotation was used to estimate the relevance of features at certain points in the music recording. During the Dagstuhl seminar, we extended this discussion by not only considering the relevance of certain feature types (e.g. representing instrumentation, harmony, rhythm, or tempo), but also the relevance of different segmentation principles based on repetition, homogeneity, and novelty. What are the musical cues used for deriving a specific segmentation boundary? Is there an unexpected musical event or a sudden change in tempo or harmony? Did the listener recognize a repeating section or a familiar phrase? Finding answers to such questions may help better understand what one may expect from automated methods and how to make computer-based approaches adaptive to account for a wide range of different segmentation cues.

References

1 Meinard Müller: Fundamentals of Music Processing. Springer Verlag, 2015.
2 Jouni Paulus, Meinard Müller, Anssi Klapuri: Audio-based Music Structure Analysis. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), 2010, pp. 625–636.
3 Jordan B. L. Smith, J. Ashley Burgoyne, Ichiro Fujinaga, David De Roure, J. Stephen Downie: Design and creation of a large-scale database of structural annotations. In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 2011, pp. 555–560.
4 Jordan Smith, Elaine Chew: Using quadratic programming to estimate feature relevance in structural analyses of music. In Proceedings of the ACM International Conference on Multimedia, 2013, pp. 113–122.

4.7 Approaching the Ambiguity Problem of Computational Structure Segmentation

Oriol Nieto (Pandora, US)

The identification of music segment boundaries has shown to be ambiguous; two subjects might disagree when annotating the same piece [1, 7]. This exposes a significant problem when developing computational approaches, which tend to be evaluated against references composed of a single annotation per track. These inadequately called “ground-truth” annotations will likely yield spurious results as long as they do not capture the inherent ambiguity of the given task.

In this seminar we discussed various ideas to approach this ambiguity problem:

- To make use of as many human annotations as possible when evaluating an algorithm's estimation for a specific music track. The SALAMI [7] and SPAM [5] datasets already contain multiple annotations for each piece.
- To weight each boundary based on a confidence value. It has been shown that humans generally agree when stating the per-boundary confidence of their annotations [1].
- To produce more than a single estimation. Let the user decide which estimation fits best her needs.
- To design and re-think current evaluation metrics using cognitive studies. Computational methods should produce estimations that are better aligned to perceptual cues. An example of this has already been published in [6].
- To annotate and estimate hierarchical boundaries. An algorithm can then be tuned to specific layers in the reference annotations. The depths of these hierarchies might differ based on the annotator, but some layers might contain a smaller amount of variations, thus reducing the ambiguity when focusing on them. Recent work towards computational hierarchical approaches can be found in[2, 3, 4].

References

1 Bruderer, M. J. (2008). Perception and Modeling of Segment Boundaries in Popular Music. PhD Thesis, Technische Universiteit Eindhoven. Retrieved from http://alexandria.tue.nl/extra2/200810559.pdf
2 McFee, B., and Ellis, D. P. W. (2014). Analyzing Song Structure with Spectral Clustering. In Proc. of the 15th International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan, pages 405–410.
3 McFee, B., and Ellis, D. P. W. (2014). Learning to Segment Songs With Ordinal Linear Discriminant Analysis. In Proc. of the 39th IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Florence, Italy, pages 5197–5201.
4 McFee, B., Nieto, O., and Bello, J. P. (2015). Hierarchical Evaluation of Music Segment Boundary Detection. In Proc. of the 16th International Society of Music Information Retrieval (ISMIR), Taipei, Taiwan, pages 406–412.
5 Nieto, O. (2015). Discovering Structure in Music: Automatic Approaches and Perceptual Evaluations. PhD Thesis, New York University.
6 Nieto, O., Farbood, M. M., Jehan, T., and Bello, J. P. (2014). Perceptual Analysis of the F-measure for Evaluating Section Boundaries in Music. In Proc. of the 15th International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan, pages 265–270.
7 Smith, J. B., Burgoyne, J. A., Fujinaga, I., De Roure, D., and Downie, J. S. (2011). Design and Creation of a Large-Scale Database of Structural Annotations. In Proc. of the 12th International Society of Music Information Retrieval (ISMIR), Miami, FL, USA, pages 555–560.

4.8 Multi-Level Temporal Structure in Music

Hélène Papadopoulos

Joint work of George Tzanetakis, Hélène Papadopoulos

Human beings process the global musical context in a holistic fashion. Music signals exhibit a complex relational structure at multiple representation levels. They convey multi-faceted and strongly interrelated information (e.g harmony, melody, metric, semantic structure), which are structured in a hierarchical way. For instance the highest-level expression of the structure (segmentation into verse/chorus, 'ABA' form etc) is dependent on musically lower-level organization such as beats and bars. Another example is that there is often a strong similarity between the chord progression of two semantically same segments.

Current computational models in MIR are limited in their capacities of capturing this complex relational structure. They usually have a relatively simple probabilistic structure and are constrained by limiting hypotheses that do not reflect the underlying complexity of music. In particular, during the Dagstuhl seminar, the problem that music analysis is typically performed only at a short time scale has been discussed. A stimulus talk on the use of semantic structure to constrain a beat tracking program has highlighted the benefit of combining longer-term analysis with shorter-term event detection (see also the abstract of Section 4.2 titled "On the Role of Long-Term Structure for the Detection of Short-Term Music Events"). How the hierarchical temporal structure of music can be described is a question that has been briefly evoked, but it remains an open discussion.

The work carried out in the emerging research area of *Statistical Relation Learning* offers very interesting ideas for modeling multi-relational and heterogeneous data with complex dependencies. In particular the framework of Markov logic networks (MLNs), which combines probability and logic, opens compelling perspectives for music processing. They seem suitable to design a multi-level description of music structure at various time scales (beat, measures, phrase, etc.) in which information specific to the various strata interact. In addition to encompassing most traditional probabilistic models (e.g. HMM), this framework allows much more flexibility for representing complex relational structure. For instance, earlier work have used structural repetitions to enhance chord estimation [1]. In a piece of music, repeated segments are often transformed up to a certain extent and present variations from one occurrence to another. Although usually strongly related, chord progressions in such repeated segments may not be exactly the same. Such variations can be accommodated by MLNs [2].

Also, among other appealing features, MLNs allow building probabilistic models that incorporate expert knowledge (constraints) in a simple and intuitive way, using logical rules. Using a language that is intuitive may be a way to make easier collaborations between musicologists and computer science people (see also the abstract of the working group "Computation and Musicology" in Section 5.8).

References

1 M. Mauch, K. Noland, and S. Dixon. *Using Musical Structure to Enhance Automatic Chord Transcription*. In Proc. of the International Society for Music Information Retrieval Conference (ISMIR), pages 231–236, 2009.

2 H. Papadopoulos and G. Tzanetakis. *Modeling Chord and Key Structure with Markov Logic*. In Proc. of the International Society for Music Information Retrieval Conference (ISMIR), pages 127–132, 2012.

4.9 The Ceres System for Optical Music Recognition

Christopher Raphael

We presented our current state of the art in optical music recognition (OMR) with a demonstration at the Dagstuhl conference.

The core recognition performed by our Ceres system system understands and represents the grammatical relationships between music notation primitives (note heads, beams, accidentals etc.) necessary for useful results. Within this context, the various recognition problems are cast as dynamic programming searches that seek the grammatically-consistent representation of an object best explaining the pixel data.

In addition to the core recognition technology, Ceres is a human interactive system in which the user guides the computer in understanding a music document. Within Ceres the user can choose candidates for symbol recognition (chords, beamed groups, slurs, etc.). After recognition the user can correct errors in two ways. In the first, the user labels individual pixels with a symbol or primitive type (beam, open note head, accidental, etc), while the system then re-recognizes subject to the user-imposed constraint. In the second, the user can change basic parameters of the recognition models, while the system re-recognizes according to the new model. For instance, we may allow or disallow augmentation dots, beams that span grand staves, two-way stems, or other possible model variations. The combination gives a flexible tool for resolving recognition problems that doesn't require any understanding of the inner workings of the system.

Our goal is to create a tool that serves as the foundation for a global effort to create large, open, symbolic music libraries. Such data are needed for digital music stands, computational musicology and many other uses. For this goal to be achieved, we must create a system where high-quality symbolic data can be created efficiently. During the Dagstuhl workshop we identified a useful partnership with Meinard Müller's group, who hope to use OMR, and the resulting symbolic music representations, to relate music scores to audio and allow search and retrieval. We look forward to pursuing this collaboration.

4.10 Musical Structure Between Music Theory, Cognition and Computational Modeling

Martin Rohrmeier (TU Dresden, DE)

The experience of music relies on a rich body of structural knowledge and the interaction of complex cognitive mechanisms of learning and processing. Even seemingly simple everyday listening experiences, such as the build up of musical tension, instantaneous recognition of a "sour" note, recognition of a variation of a familiar tune or the surprise caused by an unexpected turn of phrase, rely on a body of (largely implicit) structural knowledge. The understanding of such varieties of musical experiences lies in the interdisciplinary intersection between music theory, music cognition and computational modeling [5]. Various music-theoretical approaches proposed formalizations of high-level syntactic relations in music (e.g. [3, 4, 1]. Some of the theoretical proposals have been evaluated in experimental research (see, e.g. [2, 8]), yet the scientific potential in exploring music perception in the overlap of theory, psychology and computation remains large. During the Dagstuhl seminar, we discussed in which ways progress in complex high-level computational models of human music listening as well as in future MIR applications can be achieved by taking into account interdisciplinary insights developed in the intersection of music theory and music cognition.

References

1 Granroth-Wilding, M., Steedman, M. (2014). A robust parser-interpreter for jazz chord sequences. Journal of New Music Research, 43(4):355–374.
2 Koelsch, S., Rohrmeier, M., Torrecuso, R., Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. Proceedings of the National Academy of Sciences, 110(38):15443–15448.
3 Lerdahl, F., Jackendoff, R. (1984). A generative theory of tonal music. Cambridge, MA, MIT Press.
4 Lerdahl, F. (2001). Tonal pitch space. Oxford University Press.
5 Pearce, M., Rohrmeier, M. (2012). Music cognition and the cognitive sciences. Topics in cognitive science, 4(4):468–484.
6 Pearce, M. T., Wiggins, G. A. (2012). Auditory expectation: The information dynamics of music perception and cognition. Topics in cognitive science, 4(4): 625–652.
7 Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. Journal of Mathematics and Music, 5(1):35–53.
8 Rohrmeier, M., Rebuschat, P. (2012). Implicit learning and acquisition of music. Topics in cognitive science, 4(4), pp. 525–553.
9 Steedman, M. J. (1996). The Blues and the Abstract Truth: Music and Mentals Models, in: A. Garnham, J. Oakhill (eds.), Mental Models in Cognitive Science (Erlbaum, Mahwah, NJ, 1996).

4.11 Accessing Temporal Information in Classical Music Audio Recordings

Christof Weiß, Meinard Müller (Universität Erlangen-Nürnberg, DE)

Joint work of Vlora Arifi-Müller, Thomas Prätzlich, Rainer Kleinertz, Christof Weiß, Meinard Müller

Music collections often comprise documents of various types and formats including text, symbolic data, audio, image, and video. For example, in one of our projects, we are dealing with operas by Richard Wagner, where one has different versions of musical scores, libretti, and audio recordings. When exploring and analyzing the various kinds of information sources, the identification and establishment of semantic relationships across the different music representations becomes an important issue. For example, when listening to a performance given as CD recording, time positions are typically indicated in terms of physical units such as seconds. On the other hand, when reading a musical score, positions are typically specified in terms of musical units such as measures. Knowing the measure positions in a given music recording not only simplifies access and navigation, but also allows for transferring annotations from the sheet music to the audio domain (and vice versa). In our Wagner project, we have started to annotate measure positions within various performances for the opera cycle "Der Ring des Nibelungen" either supplied by human annotators or generated by automated music synchronization techniques. Surprisingly, even the manually generated annotations (not to speak of the annotations obtained by automated methods) often deviate significantly.

At the Dagstuhl seminar, we presented this scenario and reported typical problems. In particular, we discussed why and to which extent the task of identifying measure positions in performed music is ambiguous and what the (musical) reasons for highly deviating measure annotations are. Furthermore, we raised the question how one should evaluate automated procedures in the case where human annotators disagree. Over the course of the seminar, we found similar problems in the work of other participants. In particular, the issue of ambiguity when determining structural relevant time positions was an ongoing topic throughout the seminar.

References

1 Vlora Arifi, Michael Clausen, Frank Kurth, Meinard Müller. *Synchronization of Music Data in Score-, MIDI- and PCM-Format.* In Walter B. Hewlett and Eleanor Selfridge-Fields (ed.), MIT Press, 13:9–33, 2004.
2 Meinard Müller. *Fundamentals of Music Processing.* Springer Verlag, 2015.

5 Working Groups

5.1 Human-in-the-Loop for Music Structure Analysis

Participants of Dagstuhl Seminar 16092

Most of prior research on music structure analysis can be found in the realm of popular music. Typically, one starts with an audio recording which is then split up into segments by the system based on principles such as repetition, novelty, and homogeneity [5, 6]. Each part may then be labeled with a meaningful description such as intro/verse/chorus or a symbol 'A' as shown in the AABA form. Although the perception of structures is highly subjective, a tremendous amount of work has been proposed on developing segmentation algorithms for music structure analysis, and more importantly, on the evaluation metrics for such algorithms [9, 5, 4, 7]. However, tasks such as analyzing formal structure in tonal music require musically trained experts, who often disagree with each other in their annotations because of the ambiguity or complexity in music compositions [2]. In such cases, interactive systems that incorporate humans in the loop [8] or that are based on active learning approaches seem promising.

In this working group discussion, we shared best practices for collecting data from participants [3, 1]. We also discussed the importance of asking suitable questions to obtain meaningful answers relevant for the desired structure analysis task. Obtaining a large amount of high-quality annotated data from trained and motivated participants remains a challenge for all disciplines. However, such data, enhanced by multiple structure annotations, is especially important for MIR due to structural ambiguities (i.e., natural variations) in music. Also, multiple annotations allow for eliminating incorrect or clarifying flawed annotations. In summary, framing a question from the perspective of musicology and music cognition with quantifiable measures is one key for music structure analysis. Also, targeting an application and understanding its users may render the difficult evaluation task more feasible.

References

1 Alessio Bazzica, Cynthia C. S. Liem, and Alan Hanjalic. *On detecting the playing/non-playing activity of musicians in symphonic music videos.* Computer Vision and Image Understanding, 144(C), pages 188–204, 2016.

2 Ching-Hua Chuan and Elaine Chew. *Creating Ground Truth for Audio Key Finding: When the Title Key May Not Be the Key.* In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Porto, Portugal, pages 247–252, 2012.

3 Masataka Goto, Jun Ogata, Kazuyoshi Yoshii, Hiromasa Fujihara, Matthias Mauch, and Tomoyasu Nakano. *PodCastle and Songle: Crowdsourcing-Based Web Services for Retrieval and Browsing of Speech and Music Content.* In CrowdSearch, pages 36–41, 2012.

4 Brian McFee, Oriol Nieto and Juan P. Bello. *Hierarchical Evaluation of Segment Boundary Detection.* Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Málaga, Spain, pages 406–412, 2015.

5 Meinard Müller. Music Structure Analysis. In Fundamentals of Music Processing, chapter 4, pages 167–236, Springer Verlag, 2015.

6 Jouni Paulus, Meinard Müller, and Anssi Klapuri. Audio-based Music Structure Analysis. In Proc. of the International Conference on Music Information Retrieval (ISMIR), Utrecht, The Netherlands, pages 625–636, 2010.

7 Marcelo Rodrıguez-López and Anja Volk. *On the Evaluation of Automatic Segment Boundary Detection.* Proceedings of the International Symposium on Computer Music Multidiscplinary Research, Plymouth, UK, June, 2015.

8 Markus Schedl, Sebastian Stober, Emilia Gómez, Nicola Orio, and Cynthia C. S. Liem. *User-Aware Music Retrieval.* In Meinard Müller, Masataka Goto, and Markus Schedl, editors, *Multimodal Music Processing*, Dagstuhl Follow-Ups, volume 3, pages 135–156, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2012. http://www.dagstuhl.de/dagpub/978-3-939897-37-8

9 Jordan Smith and Meinard Müller. *Music Structure Analysis.* Tutorial at the International Society for Music Information Retrieval Conference (ISMIR), 2014, http://www.terasoft.com.tw/conf/ismir2014/tutorialschedule.html.

5.2 Computational Methods

Participants of Dagstuhl Seminar 16092

This working group discussion began by examining the performance gap between human annotators of musical structure and computational models. One key question was how to close this gap primarily by leveraging what is in the signal while recognizing that some important information used by human annotators may not be there. To this end, two different approaches were discussed in detail – one which already exists and the other which was more abstract.

- The deep neural network approach of Grill and Schlüter at OFAI [1]. The group questioned whether this highly effective system was being optimized according to the best evaluation metric, and posed a secondary hypothetical question: Could such a deep architecture learn wrongly-labeled structural boundaries?
- Theoretical large scale graphical model that extracts "everything." This theoretical approach was discussed based on evidence that some tasks which are trained across two

simultaneous (but related) tasks can outperform those trained only on one. However, the importance of defining precisely which tasks should be estimated in parallel (e.g., key, structure, chords, beats, downbeats, onsets, and so on) was considered essential.

The working group recognized the extremely high computationally demands required to run such a model, and the unresolved issue of how to train it – in particular whether all the different layers really would interact with each other equally. The group considered how key and beat information are not strongly related, but that keys are linked with chords, and beats can be associated with chord changes. A further issue concerned how to account for the fact that not all dimensions of interest might be present in the signal being analyzed. Hence, such a system would be ineffective without a measure of confidence or salience in relation to the presence or absence of such dimensions.

In summary, the group determined that it was not fruitful to attempt to build a huge model of everything musical (in this sense, it could be considered an anti-grand challenge for computational music structure analysis). Instead, the most promising next step in relation to computational models should be to identify cases where there is a meaningful and well-defined interaction between different sources of information, such as the approaches by Holzapfel et al. [2] (jointly considering beat, tempo, and rhythm), Dannenberg [3] (combining beat tracking and structural alignment), or Papadopoulos et al. [4] (jointly estimating chords and downbeats). Finally, the group revisited the issue of evaluation and how the calculated accuracy scores could be revisited in order to optimize new computational methods to the most perceptually valid measures of performance.

References

1 Thomas Grill and Jan Schlüter. *Structural Segmentation with Convolutational Neural Networks MIREX Submission*. MIREX abstract, 2015.
2 Andre Holzapfel, Florian Krebs, and Ajay Srinivasamurthy. *Tracking the "Odd": Meter Inference in a Culturally Diverse Music Corpus*. In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan, pages 425–430, 2014.
3 R. Dannenberg. *Toward Automated Holistic Beat Tracking, Music Analysis and Understanding*. In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), London, UK, pages 366–373, 2005.
4 Hélène Papadopoulos and Geoffroy Peeters. *Joint Estimation of Chords and Downbeats from an Audio Signal*. IEEE Trans. Audio, Speech and Language Processing 19(1):138–152, 2011.

5.3 Applications of Music Structure Analysis

Participants of Dagstuhl Seminar 16092

This working group focused on discussing application scenarios for music structure analysis. We began with an attempt to categorize existing applications into more scientifically and practically oriented ones. The group identified a number of applications of practical relevance in areas such as music appreciation and interaction, automatic music generation, music education and teaching, music repurposing, video production, as well as toys and games. Further examples of commercial applications using music structure analysis are, for example, *Adobe Premiere Clip* and *Jukedeck*. As for the scientific applications, the group concluded that music structure analysis is essential to reduce the complexity of music, test theories of perception, music generation, and musicology in general. The idea was brought up that, instead of struggling to reveal structural elements derived from existing music theory, music structure analysis could be applied to infer propositions for a music theory in contexts where such a theory has not yet been developed. Especially for oral music traditions this might be a promising direction for future research. Diverse findings from practical experiences were brought up throughout the discussion. For example, in the automatic creation of mash-ups, downbeats are the most important kind of structural information. The problem of recruiting and keeping users interested in the evaluation of music structure analysis results expanded upon for the case of the PHENICX project, see [1].

References

1 Cynthia Liem, Emilia Gómez, and Markus Schedl *PHENICX: Innovating the classical music experience*. Proceedings of the International Conference on Multimedia Expo Workshops (ICMEW), 2015

5.4 Rhythm in Music Structure Analysis

Participants of Dagstuhl Seminar 16092

This working group discussed issues that relate rhythm analysis to music structure analysis. On the one hand, there is a multitude of MIR tasks that involve rhythm such as beat, downbeat and tempo tracking, microtiming analysis, onset detection, and time signature estimation. On the other hand, conventional approaches to music structure analysis are often based solely on timbre and harmony, but neglect aspects related to rhythm. Recently, significant performance improvements have been achieved in tempo and meter tracking even for difficult music with soft onsets and varying tempo [2]. However, from a musicological point of view, rhythm has too often been studied in isolation from related concepts. The group discussed and identified a number of grand challenges and open issues related to rhythm analysis.

- The importance of microtiming for groove styles and short-term temporal structure [3, 1].
- Interaction between rhythm and other musical properties relevant for musical structures.
- Tracking of structures above the measure level (supra-metrical levels).
- Robust downbeat tracking.
- Exploiting larger scale repetitions for analyzing rhythmic patterns.
- Discrimination of expressive timing from bad timing.
- Visualization of conducting styles and gestures that shape the tempo.
- Identification of metrical levels that go into body movement in dance music.

References

1 Martin Pfleiderer. *Rhythmus: Psychologische, Theoretische und Stilanalytische Aspekte Populärer Musik.* Transcript, 2006.

2 Florian Krebs and Sebastian Böck and Gerhard Widmer, *An Efficient State-Space Model for Joint Tempo and Meter Tracking.* Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Málaga, Spain, October, pages 72–78, 2015.

3 Richard Ashley *Grammars for Funk Drumming: Symbolic and Motor-Spatial Aspects.* Abstract in Proc. Cognitive Science Society, Portland, Oregon, 2010.

5.5 Similarity in Music Structure Analysis

Participants of Dagstuhl Seminar 16092

In this working group, we discussed the importance of music similarity. Early on, we agreed that the measurement of music similarity by computational means is the key ingredient for music structure analysis. Unfortunately, there is no general definition of similarity readily available, as the concept depends on the signal's properties, the application context, and the user background. Usually, corpora need to be defined that reflect similarity ratings. Under the assumption that there is a common understanding of similarity among composers, performers and listeners, one then tries to measure similarity in a linear continuum. Besides the fact that measuring similarity of symbolic music representations is still an open issue, even more intricate problems arise from other music traditions that are strongly based on learning music by imitation. This was made clear by some examples from the Dunya Makam corpus [1]. While for this Turkish music tradition, similarity based on discrete pitches is reasonable, in Chinese and Indian music traditions the performance characteristics (such as inflections and vibrato) are much more important. Moreover, even though music segments may be similar from a semantic viewpoint, they often exhibit significant spatial or temporal differences. Therefore, we pointed out the importance of the design of musically meaningful mid-level features so that the subsequent similarity rating can be handled very efficiently using simple metrics (basically inner products and correlation measures).

References

1 Holzapfel, A., Şimşekli U., Şentürk S., and Cemgil A. T. *Section-level Modeling of Musical Audio for Linking Performances to Scores in Turkish Makam Music*. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 141–145, Brisbane, Australia, 2015

5.6 Structure in Music Composition

Participants of Dagstuhl Seminar 16092

Nobody really listens to generated music. We believe this is due to the fact that computer-generated music typically lacks long-term structure. Systems employing Markov models usually only look at transition probabilities on a very local level. Context-free grammars offer a step-up, but have not been very successful in constraining global structure. These days, many believe in deep learning and speak of the potential of convolutional neural nets for generating music with structure; we have not yet seen convincing results.

Since music generation can be viewed as the flip side of music analysis (as stated by one of the participants), we can use the information retrieved from the structural analysis to facilitate the music generation process. We therefore believe that generation should be a two-step process: generate the structure first (i.e., a narrative that the music follows), then generate the music based on that structure. A possible approach could be to use a tension profile model to capture structure. Such models [1, 2] can capture a lot more than people realize and listeners recognize it. This approach could potentially be combined with structure captured by information content. These two types of structure are closely related, but do not always have a linear relationship. In the approach followed by Herremans in the MorpheuS project (http://dorienherremans.com/MorpheuS), generating music that fits a tonal tension profile is combined with enforcing a higher level, top-down structure defined by a pattern detection algorithm. In conclusion, generating music into a structure allows us to overcome the problem that we have long been facing in the field of automated composition. We propose that tension and information content might be suitable structural profiles, especially when combined with a larger top-down structure.

References

1 Farbood, M. (2012). *A parametric, temporal model of musical tension*. Music Perception 29(4):387–428.

2 Herremans, D., and Chew, E. (2016). *Tension ribbons: Quantifying and visualising tonal tension*. Second International Conference on Technologies for Music Notation and Representation (TENOR). Cambridge, UK.

5.7 Structure Analysis and Music Cognition

Participants of Dagstuhl Seminar 16092

For the discussion in this session two different directions were proposed. First, the question was posed as to what aspects of the cognitive sciences and music psychology can be helpful for computational approaches to structure analysis. Second, the question on how cognitive and psychological studies can be advanced with the help of computational analysis of musical structure was raised. David Temperley, in his keynote talk at the CogMIR Conference 2013, outlined some thoughts about the relation between music cognition and music information retrieval (MIR). He stated that, for well-defined problems, MIR can work alone; for many ill-defined problems, however, MIR can profit from collaborations with the cognitive sciences. Examples that were discussed in the working group session were:

- In MIR, one rarely encounters the concept of reduction as it has been proposed in theories such as the Generative Theory for Tonal Music (GTTM) by Lerdahl and Jackendoff [1].
- Learned signal representations derived from deep learning could profit from incorporating perceptually motivated low-level features.
- Mechanisms that help humans to predict music events and shape their expectations about what is going to happen next [2] should be better exploited.

In summary, most participants of our working group agreed that MIR approaches to structure analysis should incorporate findings from the cognitive sciences related to the aspects that help human listeners chunk a piece of music into memorable segments. These processes are mainly those of expectation, reduction, and tension. However, an important problem that such an interdisciplinary exchange will face is the fact that most models for these processes have been derived and documented using notated representations of music. Making these models work with audio recordings is a major challenge that impedes straightforward incorporation of the models into MIR approaches.

References

1 Lerdahl, F. and Jackendoff, R. (1983). *A generative theory of tonal music*. MIT Press Cambridge.
2 Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

5.8 Computation and Musicology

Participants of Dagstuhl Seminar 16092

At the Dagstuhl Seminar, participants representing musicology and the computer/engineering sciences identified the hurdles that prevent more widespread collaboration between these two disciplines. In addition to their different philosophical orientations, the two fields have very different cultures. Musicological research privileges sole-authored, narrowly focused, hermeneutical, score-centered, and (sometimes) strongly opinionated writings. Computational research is multi-authored, broadly focused, descriptive, audio-centered, and leaves interpretation to the reader. Still, we also identified connecting points such as our shared interest in normal practices of musicians as well as our shared interest in analytical reduction and classification.

5.9 Music Structure Annotation

Participants of Dagstuhl Seminar 16092

In this working group, we discussed various aspects regarding the generation, the usage, and the availability of structure annotations for music recordings. The following list summarizes the main points of our discussion.

- Several systems and methodologies have been proposed to annotate the music structure of an audio soundtrack. This is not a problem in itself but should be kept in mind when analyzing the results of an algorithm on a given test-set, since the evaluation results may significantly depend on the system and methodology used to create the annotations of the test-set.
- For a single system and methodology, the annotation may also vary from one annotator to another. We therefore need to collect the information on why an annotator did his or her annotation.
- Various kinds of (known and unknown) ambiguities as well as subjectivity are sources of annotation variations.
- There seem to be a tendency for annotators to agree on large-scale structures, but not necessarily on small-scale structures.
- Annotations should be considered as a sparse graph (lattice of hypotheses), rather than a sequence of segmentation boundaries.

- The ambiguity as well as the various annotation systems and methodologies should be handled explicitly rather than being "swept under the carpet." The best solution is to collect several annotations representing the ambiguities.
- The following corpora of annotations have been identified and discussed.
  - Public: SALAMI (1400), IRISA (380), RWC (200), Isophonics (360)
  - Private (audio non-public): IRCAM (300), TUT-Paulus (600), Billboard (1000), CompMusic, ACM-Multimedia
- We should produce a document listing the various corpora and explaining their corresponding annotation systems and guidelines. It would be interesting to compare annotations of the same track resulting from the various systems.
- It is easier to detect the beginning of a segment than its end. We need a format that would easily allow for the representing of fuzzy endings. Maybe the JAMS (JSON-based music annotation) format could be used for storage [3].

References

1 G. Peeters and E. Deruty. *Is music structure annotation multi-dimensional? A proposal for robust local music annotation.* In Proc. of the International Workshop on Learning the Semantics of Audio Signals (LSAS), Graz, Austria, 2009.

2 F. Bimbot, E. Deruty, S. Gabriel, and E. Vincent. *Methodology and resources for the structural segmentation of music pieces into autonomous and comparable blocks.* In Proc. of the International Society for Music Information Retrieval (ISMIR), Miami, Florida, USA, pages 287–292, 2011.

3 E. Humphrey, J. Salamon, O. Nieto, J. Forsyth, R. Bittner, and J. Bello. *JAMS: A JSON Annotated Music Specification for Reproducible MIR Research.* In Proc. of the International Society for Music Information Retrieval (ISMIR), Taipei, Taiwan, pages 591–596, 2014.

4 J. B. L. Smith, J. A. Burgoyne, I. Fujinaga, D. De Roure, and J. S. Downie. *Design and creation of a large-scale database of structural annotations.* In Proc. of the International Society for Music Information Retrieval (ISMIR), Miami, Florida, USA, pages 555–560, 2011.

Participants

- Stefan Balke
Univ. Erlangen-Nürnberg, DE
- Juan Pablo Bello
New York University, US
- Frédéric Bimbot
IRISA – Rennes, FR
- Carmine Emanuele Cella
ENS – Paris, FR
- Elaine Chew
Queen Mary University of London, GB
- Ching-Hua Chuan
University of North Florida, US
- Roger B. Dannenberg
Carnegie Mellon University, US
- Matthew Davies
INESC TEC – Porto, PT
- Christian Dittmar
Univ. Erlangen-Nürnberg, DE
- Sebastian Ewert
Queen Mary University of London, GB
- Mary Farbood
New York University, US
- Masataka Goto
AIST – Tsukuba, JP
- Dorien Herremans
Queen Mary University of London, GB
- Andre Holzapfel
OFAI-Wien, AT
- Rainer Kleinertz
Universität des Saarlandes, DE
- Frank Kurth
Fraunhofer FKIE – Wachtberg, DE
- Cynthia C. S. Liem
TU Delft, NL
- Brian McFee
New York University, US
- Meinard Müller
Univ. Erlangen-Nürnberg, DE
- Oriol Nieto
Pandora, US
- Mitchell Ohriner
Shenandoah University – Winchester, US
- Hélène Papadopoulos
L2S – Gif sur Yvette, FR
- Geoffroy Peeters
UMR STMS – Paris, FR
- Christopher Raphael
Indiana University – Bloomington, US
- Martin Rohrmeier
TU Dresden, DE
- Mark Sandler
Queen Mary University of London, GB
- Xavier Serra
UPF – Barcelona, ES
- Jordan Smith
AIST – Tsukuba, JP
- Anja Volk
Utrecht University, NL
- Christof Weiß
Univ. Erlangen-Nürnberg, DE
- Geraint A. Wiggins
Queen Mary University of London, GB