DAGSTUHL
REPORTS

Volume 1, Issue 11, November 2011

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/dagrep

Publication date
March, 2012

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.1.11.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

www.dagstuhl.de/dagrep

Report from Dagstuhl Seminar 11441

Science and Engineering of Cyber-Physical Systems

Edited by

Holger Giese[1], Bernhard Rumpe[2], Bernhard Schätz[3], and Janos Sztipanovits[4]

1 **Hasso-Plattner-Institut – Potsdam, DE,** holger.giese@hpi.uni-potsdam.de
2 **RWTH Aachen, DE,** rumpe@se-rwth.de
3 **fortiss GmbH – München, DE,** schaetz@fortiss.org
4 **Vanderbilt University, US,** janos.sztipanovits@vanderbilt.edu

Abstract

Today, a new category of engineering systems is emerging that combines the physical with the computational in a holistic way: Cyber-physical systems (CPS). The key property of these systems is that functionality and salient system properties are emerging from an intensive interaction of physical and computational components. Traditional separation along engineering disciplines in the design of such systems leads to various quality, maintainability and evolutionary problems, and integrated theories and engineering techniques are urgently needed. The purpose of the seminar is to bring together researchers from the field, from both academia and industry to discuss the new scientific foundations and engineering principles for the vastly emerging field of CPS.

Seminar 01.–04. November, 2011 – www.dagstuhl.de/11441
1998 ACM Subject Classification C.0 [Computer Systems Organization] General
Keywords and phrases Embedded systems, real-time systems, control, composition, system integration, design automation, model-driven development, validation & verification
Digital Object Identifier 10.4230/DagRep.1.11.1

1 Executive Summary

Holger Giese
Bernhard Rumpe
Bernhard Schätz
Janos Sztipanovits

Today, a new category of engineering systems is emerging that combines physical processes with computational control in a holistic way: Cyber-physical systems (CPS). Cyber Physical Systems are engineered systems of synergistically interacting physical and computational components. The key property of these systems is that functionality and salient system properties are emerging from an intensive interaction of physical and computational components. As the computational components are aware of their physical context, they are intrinsically distributed, (time)-synchronizing, have to cope with uncertainty of sensoric-input and need to produce real-time reactions. Consider an unmanned aerial vehicle (UAV) with active wings. In such an UAV, a cyber-physical system may consist of an embedded controller, monitoring the airflow over the wing surface and electromechanical actuators modulating the airflow to ensure laminar flow such that the vehicle is capable of extreme maneuvers. Unlike more traditional embedded systems, full-fledged CPSs are often designed as networks of interacting

elements including autonomous automotive systems, medical monitoring, process control systems, distributed robotics, and automatic pilot avionics. The question naturally arises: are cyber-physical systems fundamentally different such that they need a different fundamental science, a different development approach, or is the current approach sufficient and no new research is necessary? We argue that new science, new techniques, and a new view are necessary. Traditional separation along engineering disciplines in the design of such systems leads to various quality, maintainability and evolutionary problems: thus, integrated theories and engineering techniques are urgently needed. The technology is pervasive, transcends industrial sectors and serves as the engine of innovation for new generation of products. CPS is also a disruptive technology that transforms established industries, may create new ones and possibly rearranges the status quo of development in entire industrial sectors. Current industrial experience tells us that we have reached the limits of our knowledge regarding integration of computers and physical systems. These shortcomings range from technical limitations in the scientific foundations of cyber-physical systems through the engineering processes to the way we educate engineers and scientists that support cyber-physical system design. However, besides the National Science Foundation initiative in the US, the topic is currently addressed by initiatives such as intelligent and autonomic automobiles, ambient intelligence, self-organizing embedded systems, plant-control and reparation, self-optimizing mechatronic systems, 'smart' power grids, in-home medical assistance devices, etc. This seminar focused on the scientific foundations and the engineering aspects of cyber-physical systems by bringing together researchers from both academia and industry to discuss the new scientific foundations and engineering principles for the vastly emerging field of CPS.

2 Table of Contents

3 Overview of Talks

3.1 A Network-centric Perspective on Cyber-Physical Systems

Luis Almeida (Distributed and Real-Time Embedded Systems Lab (DaRTES), Instituto de Telecomunicações, University of Porto, PT)

URL http://www.it.up.pt

Cyber-Physical Systems (CPS) present a unified view of computing systems that interact strongly with their physical environment, from typical embedded systems to networked monitoring and control, ubiquitous systems, systems of systems, etc. A common feature of almost all CPS is that they heavily rely on networking. Therefore, the network plays a central role in supporting the needed system-wide properties, being timeliness a particularly important one as dictated by the dynamics of the associated physical process. However, a generalized approach to provide real-time communication for CPS is lacking. There is a well known body of work towards latency-constrained communication within distributed embedded systems, which have a clear infrastructure and requirements, but the same is not true in processes that are distributed over large areas, possibly relying on the Internet, where the infrastructure is largely unknown, and the network has essentially been dominated by throughput and scalability, with timeliness being a second concern.

We claim that new CPS applications, such as Smart-Grids, Remote Interaction, Collaborative Robotics, etc, require openness together with tighter timeliness guarantees that can only be achieved with a paradigm shift from packet switching with class-based scheduling to channel reservation-based communication. We define the challenge and state some of the directions that will potentially provide scalable and open latency-constrained communication.

We end the presentation with a brief reference to our recent work towards that goal, based on scaling previous work on flexible and composable approaches to real-time communication for distributed embedded systems.

3.2 Extending Passivity to Guarantee Properties in CPS Design

Panos J. Antsaklis (University of Notre Dame, US)

Joint work of Antsaklis, Panos J.; Gupta, Vijay; Goodwine, Bill
URL http://www.nd.edu/~pantsakl/Publications/PublicationsListing.html

In Cyber-Physical Systems large number of heterogeneous cyber and physical subsystems are networked, interacting tightly, may change dynamically and may expand or contract. Designing and preserving properties of a CPS over its lifespan is very challenging. Passivity and dissipativity are energy like concepts that offer great promise in guaranteeing properties, such as stability, in complex heterogeneous interconnected systems that are changing dynamically. Passivity indices that provide a measure of the degree of passivity are used to generalize classical results in interconnected systems, and results for continuous, discrete and switched systems in networks with delays, event triggered architectures, conic systems and systems with symmetries are shown.

3.3 CPS from a Control Perspective

Karl-Erik Arzen (Lund University, SE)

The term Cyber-Physical Systems is used to denote applications where a tight integration is required between the computing parts of the application and the physical parts of the application. However, this is the normal case in control engineering. The topic of this talk is what distinguishes CPS from classical control. When designing complex artifacts and systems separation of concerns is a good design principle. However, the current focus on resource efficiency is cross-cutting and requires solutions based on integration and co-design. Whether a control application should be considered CPS or not, in my mind depends on (at least) three different items. A control system is CPS when temporal effects of the implementation platform caused by computing and communication, needs to modeled and included in the design at a more detailed levels than what is traditionally done in computer-based control.

Second, control applications of a CPS nature are typically more distributed and decentralized in nature than the classical, more centralized, control approaches. Third, a control application can be considered to be CPS when the system under control itself is a computing and/or communication systems, e.g., a data center on an embedded MPSoC.

In this talk the work on co-design of control and computing system at Lund University will be presented. The talk presents the Jitterbug toolbox for analyzing how temporal non-determinism effects control performance, the jitter margin that gives analytical bounds on how much jitter in sampling and actuation a controller can tolerate, the TrueTime simulator for CPS control systems, and recent results on event-based and sporadic control. The presentation also briefly touches upon the work being done in Lund on distributed control, including distributed convex optimization, distributed Model Predictive Control, and distributed control of positive systems.

3.4 Towards Verifying CPS with Structural Dynamism

Basil Becker, Holger Giese (Hasso-Plattner-Institut – Potsdam, DE)

In our work we focus on distributed, decentralized and safety-critical cyber-physical systems where the combination of networking and control results in new opportunities. We especially emphasize systems which face inherent complex structural dynamism due to such phenomena as mechanical coupling of moving parts, mobility and coalition building or self-organization. Thus, a technique that aims at verifying such systems first has to cope with the complexity introduced by the system's physical nature and second has to be able to cope with complex structural changes that also impact the physical behavior.

For our approach we employ graph transformation systems with continuous behavior in form of ODE to capture the behavior of such CPS. Furthermore, we extended our invariant checking technique – a technique that can statically verify inductive invariants for sets of timed graph transformation rules and timed graph constraints – to also deal with differential equations for the continuous behavior. We exemplify our approach using a system of

autonomous shuttles. For these shuttles we want to verify that platooning is safe and no collision occur.

Obviously in a real-world system a tremendous number of situations exist where a collision could happen. We use graph transformation rules to describe the shuttles abstract movement on the topology and the creation of a platoon and graph pattern to describe collisions that have to be excluded. Continuous attributes, whose derivation is given through ODE, together with attribute constraints describe speed, acceleration and position.

3.5 CPS and Multi Paradigm Modeling in ModHel'X

Cecile Hardebolle, Frederic Boulanger (Supélec - Gif-sur-Yvette, FR)

Joint work of Boulanger, Frédéric; Hardebolle, Cécile; Jacquet, Christophe; Marcadet, Dominique
Main reference Frédéric Boulanger, Cécile Hardebolle, Christophe Jacquet, Dominique Marcadet, "Semantic Adaptation for Models of Computation, " in Proc. of ACSD 2011, IEEE Computer Society, pp. 153–162.
URL http://dx.doi.org/10.1109/ACSD.2011.17

Cyber Physical Systems deal with a mix of software-based and physical components. We consider that the design of such systems raises two challenges:
(a) the use of heterogeneous modeling paradigms for designing components and
(b) the composition of the models of components, which obey different modeling paradigms, in order to be able to reason globally on a CPS under design.
We present ModHel'X, an experimental platform for multi-paradigm modeling and simulation.

Through the example of a car power window, we illustrate our approach of the representation of modeling paradigms in a form that facilitates the composition of models. Then, we present the key concept of semantic adaptation, which defines explicitly how models that use different modeling paradigms are composed to build a global heterogeneous model. We illustrate on the power window example how semantic adaptation can be decomposed along three axis: the adaptation of data, the adaptation of time notions and the adaptation of control flow.

We also show the benefits of modeling the adaptation explicitly and apart from the models. We conclude with an overview of our current research directions on multi-paradigm modeling.

3.6 CHROMOSOME: Building blocks for CPS platforms

Christian Buckl (fortiss GmbH – München, DE)

This talk interprets Cyber-Physical Systems (CPS) as systems consisting of a set of subsystems, that were developed independent of each other, and that interacts with the environment. Each subsystem can fulfill its main functionality independent of each other. Through integration, further / better functionality can be achieved.

Based on this definition, one major challenge of CPS is the integration of heterogeneous subsystems. In the past, integration was solved by the use of domain-specific middleware.

Due to the cross-domain nature of CPS, a middleware-based solution for integration must also support requirements from different domains. More specifically, the solution must both satisfy requirements coming from the embedded domain such as predictability and safety and requirements coming from the internet domain such as adaptivity and plug&play capability.

The talk presents the CHROMOSOME middleware, a middleware directly targeted to CPS with the goal to solve above mentioned research questions. The main properties of the solution are discussed and a set of applications that are developed using the middleware is described.

3.7 Co-modelling and Co-simulation for Dependable Cyber-Physical Systems

John S. Fitzgerald (Newcastle University, GB)

The effective use of model-based formal methods in the development of dependable cyber-physical systems requires the integration of discrete- event models of software-rich elements such as controllers, with heterogeneous, often continuous-time, models of their environments. We discuss an environment for collaborative modelling and co-simulation in which a reconciled operational semantics of two formalisms provides a basis for early-stage examination of design alternatives. The approach has been realised using the VDM and 20-sim formalisms implemented in their respective simulation tools. We briefly consider the structure of the tools, the modelling of errors, and of error detection and recovery mechanisms using both discrete and continuous sides of the co-simulation. We discuss the provision of libraries of patterns for fault and fault-tolerance modelling in this context, the need to provide support for collaboration, and the potential for treating CPS as systems-of-systems.

The first group of 8 slides provides the core of the presentation. The remaining slides provide background material on the co-simulation framework, and an illustrative example based on safety kernel and voter patterns applied to paper processing machinery.

This work is primarily carried out in the FP7 DESTECS project (www.destecs.org). Future work on systems-of-systems in this context will be carried out in the FP7 COMPASS project (www.compass-research.eu).

This presentation is expected to fit with Theme 5 (design paradigms).

3.8 Multicore Platform Enablement for Cyber Physical Systems

Andreas Herkersdorf (TU München, DE)

Joint work of Herkersdorf, Andreas; Lankes, Andreas; Rauchfuss, Holm; Walla, Gregor; Zeppenfeld, Johannes;
URL http://www.lis.ei.tum.de

Cyber Physical Systems (CPS) design expands the cross-layer hardware/software co-design and co-optimization methods of traditional distributed embedded computing systems into the process specifics of different physical domains.

This is a new quality since interdisciplinary skills, methods and techniques, such as microprocessor and computer architecture, high-speed quality of service (QoS) networking, control theory, real-time computing, bio-medical and electromechanical engineering, autonomous computing, and more, have to be linked or even merged in order to achieve a holistic system optimization. Common modeling abstractions and interface semantics are critical to cope with the ever increasing complexities of these systems.

The institute of Integrated Systems at TU München has a strong research focus on architectures, design methods and tools for application-specific multicore systems on chip (MPSoC). Target application areas are Internet network processing, computer vision and driver assistance in automotive, as well as mobile robotics and mobile communications. We develop new designs and prototypes of 2D and 3D network on chip (NoC) interconnects, dedicated function hardware accelerators, hardware-supported virtualization and process synchronization techniques in order to optimize the energy efficiency, dependability, flexibility and real-time capability of scalable MPSoC platforms.

Reuse of existing MPSoC hardware and software building blocks is a key pre-requisite for developing application or customer specific solutions within reasonable time windows and with a high chance for first time success. Our research interests in CPS are related to investigating and provisioning physical domain specific hardware and hardware-aware software enhancements for scalable multicore computing platforms and corresponding augmentations to trace-based system-level exploration tools. Modified roles, types and physical realizations of interfaces between the classical compute and different physical domains are aspects I expect to obtain new insights from attending the upcoming seminar.

3.9 Analytic Virtual Integration of Cyber-Physical Systems & AADL: Challenges, Threats and Opportunities

Jerôme Hugues (ISAE – Toulouse, FR)

The design and implementation of cyber-physical systems gather multiple domains, from low-level physics up to complex control of systems to implement a full function. Such complexity requires particular strategy to characterize each level of abstractions, and then integration to ensure the system under consideration is correctly built. The advent of Model-Based Engineering is often perceived as a silver bullet to achieve all these complex tasks: the system designer can master its design through proper model artifacts (blocks, connections, properties, ...), virtual integration of system blocks, and analysis.

However, current MBE processes usually cover vertical analysis, and address only a few aspects like scheduling or behavioral analysis, while CPS would require also horizontal analysis of the system, combining analysis results.

In this position paper, we review experiments on the use of AADL to design CPS, and highlight challenges, threats and opportunities to support analytical virtual integration.

3.10 Some Issues on Formal Safety Analysis and Verification in Industrial Practice

Michaela Huhn (TU Clausthal, DE)

Joint work of Huhn, Michaela; Bessling Sara; Milius, Stefan; Daskaya, Ilays;
Main reference Ilyas Daskaya, Michaela Huhn, Stefan Milius, "Formal Safety Analysis in Industrial Practice," FMICS 2011, pp. 68–84.
URL http://dx.doi.org/10.1007/978-3-642-24431-5_7

Formal safety analysis techniques and verification enable to mathematically reason on functional safety properties of a system under design. Whereas the progress in research on verification techniques is tremendous, industries still hesitate to integrate these techniques in their quality assurance process, even in cases where design models are already available in a formally founded development framework. We investigate two industrial case studies and identify two enhancements that hopefully may pave the way for an increasing use of formal analysis techniques:

- Even with formal safety analysis and a formal model of primary faults at hand, it's not obvious how to map the fault propagation and transformation as it is described e.g. by a fault tree, into the behavioral system design on which the safety analysis may formally examine its effect.
- Even with our medium size industrial case studies we observed intense complexity problems that could not be overcome by employing different heuristics like abstraction and compositional verification.

Both case studies indicate that the modeling style has a significant impact on the complexity of the verification task. We finally succeeded to prove critical properties by combining abstraction and model transformation from SCADE to UPPAAL timed automata. Both case studies indicate found that the modeling style has a significant impact on the complexity of the verification task.

3.11 Contract-Based Design of Embedded Systems

Hardi Hungar (OFFIS – Oldenburg, DE)

This work concerns the semantical foundations for a compositional development method. The main design units in the method are components, whose nature comprises two facets: Assumptions about the environment in which they may be placed, and guarantees about their behavior, provided the assumptions are met.

Components may be described declaratively in the form of specifications, or operationally by models. A compositional notion of refinement permits to relate more precise versions of design units with previous ones. Refinement distributes over the structure of decomposition into parallel units. A more general notion of realization captures the change of levels of abstraction or the transgression from a conceptual perspective to a more concrete one, such as form a functional view to a logical or technical one. By incorporating all these concepts, this work provides the foundation for being able to express precisely in which way the final design implements the requirements which have been formulated at the start of the development process.

3.12 Polyglot: Modeling and Analysis for Multiple Statechart Formalisms

Gabor Karsai (Vanderbilt University, US)

In large programs such as NASA's Exploration Systems, multiple systems that interact in safety-critical protocols are already designed with different Statechart variants. In order to verify these safety-critical systems, a unified framework is needed based upon a formal semantics that captures the different Statechart formalisms. This paper first provides a parametric formal semantics developed in SOS that captures the common core of Statecharts with extensions for different dialects, addressing previous limitations. It then describes the architecture of our implemented unified framework, which translates Statechart models to Java, with pluggable semantics for different variants operating in a generic execution environment. This environment has been integrated with the Java Pathfinder model checker, providing analysis and verification capabilities including concrete model checking against requirements and test-vector generation. The paper outlines the application of this unified framework during requirements analysis of the launch abort protocol between the Orion capsule and the Ares launch vehicle.

3.13 Flexible Multicast Authentication for Time-Triggered Embedded Control Network Applications

Philip Koopman (Carnegie Mellon University – Pittsburgh, US)

Security for wired embedded networks is becoming a greater concern as connectivity to the outside world increases. Protocols used in these networks omit support for authenticating messages to prevent masquerade and replay attacks. The unique constraints of embedded control systems make incorporating existing multicast authentication schemes impractical. Our approach provides multicast authentication for timetriggered applications by validating truncated message authentication codes (MACs) across multiple packets.

We extend this approach to tolerate occasional invalid MACs, analyze our approach through simulated attacks, and give an upper bound on the probability of successful attack. This approach allows a tradeoff among per-packet authentication cost, application level latency, tolerance to invalid MACs, and probability of induced failure, while satisfying typical embedded system constraints.

3.14 Avoiding the Top 43 Embedded Software Risks

Philip Koopman (Carnegie Mellon University – Pittsburgh, US)

This talk briefly distills the lessons learned from almost 100 design reviews of industry embedded software projects. In brief, most critical project risks had a root cause of process problems rather than technical problems, and most risks were gaps (developers not knowing to do something) rather than incorrect execution of a desired process step.

3.15 Security of CPS: Secure Embedded Systems as a Basis

Christoph Krauss (Fraunhofer AISEC – München, DE)

URL http://www.aisec.fraunhofer.de/

The security of Cyber Physical Systems (CPS) is of paramount importance to enable many application scenarios and to achieve a broad user acceptance. In addition to communication security, the security of a used embedded system itself must be ensured since such systems are often deployed in unattended or even hostile environments which enable an adversary to manipulate or compromise these systems.

This talk presents an overview of research performed at Fraunhofer AISEC to secure embedded systems which enables a secure application in CPS. First, secure elements, which provide secure storage and execution of (cryptographic) operations, are introduced. Attacks on secure elements such as Side Channel, Probing & Forcing, and Fault Injection performed in the AISEC labs are briefly introduced to show which knowledge is required to design secure elements.

Second, this talk presents in more detail a relatively new approach to secure embedded systems called Physical Unclonable Functions (PUF). These PUFs exploit unclonable physical characteristics which enable the unique authentication of a system and provides mechanisms for system integrity.

3.16 CPS Safety

Peter Bernard Ladkin (Universität Bielefeld, DE)

I discuss some issues with the safety of the kinds of systems participants call "Cyber-Physical Systems".

Postscript: I wrote a blog post and a note using as illustration an unfortunate incident that occurred in GB on the motorway M5 the day we left. Exactly the same happened in Germany on the Autobahn A31 a week later.

References

1 Peter Bernard Ladkin. *Assurance of Cyber-Physical Systems*, 2011. http://www.abnormaldistribution.org/2011/11/17/assurance-of-cyber-physical-systems/

2 Peter Bernard Ladkin. *The Assurance of Cyber-Physical Systems: Auffahr Accidents and Rational Cognitive Model Checking*, 2011. http://www.rvs.uni-bielefeld.de/publications/Papers/20111117CPS.pdf

3.17 A Framework for Software/System Certification

Tom S. Maibaum (McMaster University – Hamilton, CA)

In this talk we will address the nature of certification in the context of critical systems, decomposing it, by means of a new philosophical framework, into four aspects: evidence, confidence, determination and certification. Our point of view is that establishing the safety (in a very general sense) of a system is a confidence building exercise much in the same vein as the scientific method; our framework serves as a setting in which we can properly understand and develop such an exercise.

3.18 Putting the physics in the design of Cyber-Physical Systems

Pieter J. Mosterman (The MathWorks Inc. – Natick, US)

In the design of Cyber-Physical Systems, physics plays a crucial role.

Models of physics at a macroscopic level often comprise differential and algebraic equations. These equations typically require computational approaches to derive solutions. Approximations introduced by the solvers that derive these solutions to a large extent determine the meaning of the models, in particular when discontinuities are included. In reasoning about models that are solved computationally it is therefore imperative to also model the solvers. This presentation shows how performance of a cyber-physical system may be affected by physics and conceptualizes the modeling of computational solvers.

Opportunities that derive from the availability of solver models are presented and a control synthesis approach for stiff hybrid dynamic systems based on model checking is outlined.

3.19 CPS in technical medicine – from training to a clinical surgical setting

Jerzy W. Rozenblit (University of Arizona – Tucson, US)

Laparoscopic surgery is a surgical technology that can minimize recovery time and postoperative pain. However, with this procedure surgeons lose many of the tactile and visual cues that they rely upon in conventional surgery. Current research and commercial products focus on virtual simulation of procedures, generation of haptic feedback for training, and automated control of the laparoscope in the operating room (OR). This talk will provide an overview of the concepts, will discuss some of the existing systems, their advantages and shortcomings. Then a design concept for a surgical training and assessment system that provides sensing and reasoning capabilities for laparoscopic surgery will be presented. The system implements sensors and offers real-time feedback capability that can enhance sensory input for surgeons. The key issues from a cyber-physical perspective such as modeling paradigms, integration, safety criticality, and real-time support of clinical procedures will be raised.

3.20 Modelling and Structuring CPS

Bernhard Rumpe (RWTH Aachen, DE)

The upcoming CPS paradigm allows us to rethink how we develop systems. Today we experience that when structuring and decomposing systems during development, the partitioning into hardware (system, electronics etc.) and software (program) is done pretty early. However, this has some drawbacks with respects to integration and reuse. A functional decomposition into reusable components and an integrated, feature based composition of the implemented components will allow us to develop with more efficiency and quality.

However, this needs new integrated forms of modelling – and modelling languages that integrate mathematical calculus and digital theory of discrete, event based system. A sound and integrated foundation is necessary to be able to analyse, synthesize or simulate functions and systems developed in such a manner. We do provide our work in progress on rethinking modelling in a structured, modular way allowing new forms of decomposition and analysis on CPS components and systems.

3.21 Integrating Engineering and Operation of CPS

Bernhard Schätz (fortiss GmbH – München, DE)

As CPS are generally large-scale and long-living systems, they are in constant evolution, rendering the classical "develop–commission-decommission"-life cycle of embedded systems inadequate. The blurring between the operation and the engineering phase requires concepts like self-documentation, self-management, and self-protection with adequate techniques like built-in reflection, built-in maintainability, and built-in robustness, effectively turning a CPS into its own IDE.

3.22 Some challenges in modelling Cyber Physical Systems

Hans Vangheluwe (McGill University – Montreal, CA)

1. The need to deal with models in different formalisms. From an expressiveness point of view, it is necessary to specify embeddings of models in one formalism into models in another formalism, as in ModHel'X (see the presentation by Cecile Hardebolle). To study multi-formalism models, co-simulation may be used, but when for example symbolic analysis is desired, model transformation (onto an appropriate formalism) is more appropriate.
2. The need for modular building blocks which encompass physical, control, and software aspects. Experience with multi-physics modelling using Modelica suggests that it is possibly to design and use modelling constructs which encapsulate these aspects. Inside the building blocks, the interactions between the aspects are modelled. Composition is done through connection of physical ports, control ports, and software (event) ports. Note that ultimately, all aspects will be reduced to computation, which should be modelled explicitly (see the presentation by Pieter Mosterman).
3. The need to extend modelling language engineering from design languages only to I/O, trace, and properties languages and their inter-relationships. his is most acute in the case of Domain-Specific Languages. The added challenge in CPS is the need to engineer new modelling languages for dynamic-structure, adaptive, context-aware systems. Such languages need to be modularly designed, containing parts to describe (1) when a change occurs, for example based on trace-matches (2) how the model structure/formalism changes, for example based on a rule-based description and (3) how to consistently (re-)initialize the new model, for example based on physical conservation laws (as in Pieter Mosterman's PhD thesis).

3.23 Certification Challenges in Cyber Physical Systems – and How to Meet Them

Alan Wassyng (McMaster University – Hamilton, CA)

Cyber physical systems are large, highly complex, interconnected, real-time computer systems embedded in a physical environment. Almost all cyber physical systems that we know about today are safety or financially critical. They have to be incredibly dependable and safe, and we need to be able to demonstrate that they are dependable and safe through some form of certification. However, when we look at the major challenges in certifying software intensive systems, we see a familiar set of items that cause us problems: large, complex, real-time, networked and distributed.

So, what do we do? One approach is to identify system properties that we hold inviolate, and then *prove* that they are never violated. What else can we do?

Mathematical verification of correctness with our current technologies is a non-starter for these very large systems. Traditional testing is probably intractable in most cyber physical systems. Perhaps the best thing we can do is look for ways in which we can reduce the complexity of the system so that existing certification approaches can be used with success. Some industries already separate safety and control systems (for instance). This is a true and

complete separation. In such cases the safety system is demonstrably less complex than the control system, and is then amenable to various certification approaches. The presentation will conclude with a discussion on some illustrative examples.

4 Working Groups

4.1 Modeling

From the previous texts it becomes clear that an new category of engineering is emerging which combines the physical with the computational in a holistic way: cyber-physical systems (CPS). The key property of these systems is that functionality and salient system properties are emerging from an intensive interaction of physical and computational components. Traditional separation along engineering disciplines in the design of such systems leads to various quality, maintainability and evolutionary problems, thus integrated theories and engineering techniques are urgently needed. The purpose of the seminar is to bring together researchers from both the academia and industry to discuss the new scientific foundations and engineering principles for the vastly emerging field of CPS.

We thus have established a number of working groups that discussed several aspects of CPS and how to approach their development, maintenance, etc. The following groups were built:

1. Modeling
2. Analysis, Verification, Validation
3. Adapting, Evolving, Operating, Platforms

Based on plenary presentations of the results as enriched by discussions well as permutations of participants in the working groups, it became clear that although all these aspects can be discussed separately, they are widely connected.

Important topics regarding modelling of cyber-physical systems were for example:

1. What is the use of models for development, configuration at runtime, maintenance and evolution of CPS? What other forms of uses CPS might have.
2. Assuming that models useful: What are the appropriate modeling languages and what is the necessary tooling infrastructure for this?

There was a common sense that existing general purpose modeling languages do have some positive impact and are of good use in many CPS-development projects. However, the language to be used has severe impact on the things that can be expressed. As it is generally agreed CPS need a new way of thinking, it is a necessary conclusion that the development of CPS needs more appropriate languages to capture individual aspects as well as integrated views of CPS. One core deficiency is the lack of a well understood integration of calculus, which is the basis for industrial process and control theory, and the digital theory of state machines, which is used in digital systems. A sound and integrated hybrid theory is essential for a fruitful development of cyber-digital and controlled-physical systems.

Based on such a theory, individual languages expressing core drivers (risk factors) of the respective domains and in particular domain specific languages are necessary. It was generally seen, that each of the many domains that belong to CPS (such as automotive, smart phones, trains, airplanes, power plants, etc.) will have their own vocabulary that needs to be expressable within the respective languages.

And of course, the modeling languages in questions are there to develop CPS as well as to analyze, verify and validate their properties. This enforces languages to be rather expressive on one side, but also to be restrictive on the other side, because this allows more analytical machinery to be applied.

4.2 Analysis, Verification, Validation

The main result of the Working Group on Analysis, Verification, Validation was that due to the transition from standard embedded systems towards cyber-physical systems many of the crucial elements for the analysis, verification and validation in that area such as standards (in particular safety), the handling of requirements, procedures and regulations for certification, means for isolation are essentially *'broken'* and do not longer work resp. cannot be employed as today.

For the area of *interfaces and composition* the group identified the challenge that new solutions that can handle unwanted coupling in the physical domain and cross-disciplinary interfaces have to be established. It was identified that freedom from interference due to containment/isolation has to be extended to cyber-physical systems. Also assumptions have to be made more explicit and must be monitored at runtime by the system rather than taken for granted at design time to contribute to some 'defense in depth'. Also better restricting interfaces such that they only expose really necessary capabilities are required which will likely require more discipline as well as more formality. An unsolved problem is, however, that known concepts for contracts do not solve non-local properties (e.g. control and stability). Probably, to overcome these obstacles a shift towards design for verification seems unavoidable.

Furthermore, concerning *emergent behavior and scale* two developments that seen contrary to each other where observed. On the one hand radical performance improvements based on non-local information from the cyber-space seem often possible and are very attractive. On the other hand this raises a number of serious problems concerning trust. For somehow cooperating agents this trust result in only *relative* safety. A common understanding of the true state is one crucial requirement in establishing trust, which is challenging. Also trust in adherence to agree upon rules may allow additional relative safe options. However, it seems that the envisioned direction has to be accompanied by means for detecting incidents and a shift towards design for (technical) accountability as otherwise the risks seem not justifiable and capabilities to mitigate problems seem not sufficient.

Cyber-physical systems more actively operate in more dynamic contexts and thus have to handle *uncertainty* that results from phenomena such as abnormal behavior, rare events, openness, or evolving structure. In this context the fact that unknown dependencies can break independence assumptions has been identified as crucial. To approach the resulting problems possible options seem to be looking at worst cases (if not too pessimistic), following analysis approaches that cover uncertainty (where feasible) as well as delaying the problem to the runtime where the concrete problem can be observed effectively resulting in the reduction of uncertainty which permits to better react most appropriately.

Finally, the requirement that cyber-physical systems support *adaptation* results in similar problems concerning the uncertainty, which result from the context (i.e., the environment) and from the system and its components itself. The structure as well as behavior may evolve due to adaptation steps and suitable approaches for analysis, verification and validation therefore would require a treatment that covers all potentially unbounded many possible

configurations that may result from such evolution as well as the different possible evolution paths. In particular, the already discussed challenges for *interfaces and composition* and concerning *emergent behavior and scale* further complicate this problem as probably not even the complete information required to describe the possible evolution is available.

4.3 CPS-Platforms

Cyber-Physical Systems are systems integrating physical and organizational processes by means of information and communication technology (as part of devices, building, vehicles, transport infrastructure, production facilities, medical/logistic/coordination/management processes) that

- via sensors and actors directly sample and influence physical processes
- process/store sampled data and (re)actively interact with the physical and digital word
- are linked via digital communication infra structure with and within global networks
- use globally available data and services
- provide a collection of dedicated multi-modal man/machine interfaces

Obviously, in contrast to classical business information systems as well as embedded systems, with such heterogeneous requirements, the concept of a platform plays a central role in a cyber-physical system, since this platform must

- allow to interface with the physical world and the digital world
- support the integration of different sub-systems
- enable integration on a large scale

Furthermore, since cyber-physical systems are in general large-scale systems and long-living systems, the platforms must cater to the needs of *operating these systems, including maintaining, updating, and evolving them.* These requirements must be supported as *built-in properties of a platform*, to effectively design, construct, and operate cyber-physical systems. In short, in a CPS the distinction between analysis, design, implementation, commissioning, operation, and decommissioning is no longer sensible. Even more pointedly, with a CPS the development environment and the operation platform amalgamate, making a CPS an IDE, operating platform, and systems at the same time.

4.3.1 Drivers

The longevity of cyber-physical systems – together with the above-mentioned resulting requirements of built-in mechanisms to operate, maintain, adapt, and evolve – leads to a set of driving forces, characterizing the capabilities of CPS platforms. These drivers include

Changing Requirements: Like classical software systems, CPS has deal with changing requirements. This includes the need to meet new demands of users to maintain user satisfaction, to deal with requirements triggered by platform/hardware evolution, as well as interoperability requirements to support the stepwise integration of systems.

CPS are repeatedly extended: Driven by either changing requirements or by availability of new technology, CPS are repeatedly enhanced and extended in their prolonged life-time.

Different life cycles of parts: Due to the (technical) heterogeneity of a CPS, parts of the system have rather different life cycles. While low-cost/COTS elements (e.g., sensors, computation platforms) tend to have rather short life cycles, high-end and individual parts (e.g., production equipment, communication infrastructure) in general have long life cycles. Software – or at least the implemented functionality – is one of the parts with the most extended life cycle.

Complex systems not defined from the beginning: Complex systems in general are not constructed in a big-bang fashion but require an incremental, or even evolutionary strategy. Furthermore, as CPS generally build on pre-existing infra-structure and are often constructed by integration of those, often there is no such thing as a master blueprint from the beginning.

Cyber + physical makes requirements engineering much harder: CPS require the integration of different disciplines (computer science, electrical engineering, mechanical engineering, etc) as well as different domains (e.g., process automation, logistics, communication), thus often requiring to form an integrated understanding of the needs of each of the participating stakeholders in a stepwise fashion.

CPS are open or used in changing environments: As CPS are integrated in open and therefore (partially) unrestricted environment, a CPS must be prepared to adapt to changes in the environment. This also intended as well as unintended use of the systems (e.g., attackers getting smarter, requiring better security mechanisms).

Need for self-x: As a CPS has to act in an open environment with untrained users, dynamically added components, or occurring faults, the system must be capable to reflect on its structure, monitor itself for its health, or actively take actions to maintain or re-establish its integrity.

All of these mentioned drivers are even strengthened in their effect due to the circumstance that many of the changes (to software, hardware, etc) must be done to the running system.

4.3.2 Challenges

To answer to these driving forces, defining the capabilities of a CPS platform, several challenges have to be addressed:

- Change can occur at different levels in a CPS, starting from the swapping of a defect sensor up to the dynamic re-integration of a system when reestablishing overall integrity. As each form of change has a different impact and requires different means of dealing with it, hierarchies of changes/adaptations/levels/classes + interdependencies are needed.
- Currently, there are no defined procedures to evolve systems w.r.t. so safety/security; more specifically, there is no certification procedure ensuring the safe and secure evolution of a CPS. To meet the public need for safety and security, such procedures have to be established.
- Currently, the impact of change is even hard to judge for classical systems. Therefore, in CPS it becomes even more complicated to understand what "direction of change" is needed to keep it dependable, compliant, safe, secure, etc.
- A CPS has to be prepared to deal with change – either by providing mechanisms to facility (manual) change or by pro-actively executing the change autonomously. Thus, techniques for the modeling of change and for change are needed.
- A large-scale system – as most CPS are – in generally cannot be shutdown for maintenance or adaption. Therefore, preserving system integrity while adapting the running system is running must be supported.
- As a CPS is not meant to be shutdown in case of problems, it is essential to keep a system viable (i.e., self-monitoring, self-healing, etc). Specifically, a CPS must provide built-in mechanisms for recovery (e.g. in case of faults or unwanted modifications).
- As a CPS may autonomously deal with necessary changes, a CPS platform must support the dynamic selection/allocation/partition from different possibilities of adaptation, analyzing tradeoffs and identifying optimal adaptations (incl. in cooperation with users, or other CPS).

4.3.3 Cross-Cutting Issues

When addressing these challenges, it becomes obvious that the integrated development, operation, maintenance, and adaption mechanisms offered by a CPS platform affect several identified cross-cutting concerns. Here, the issue of evolution/adaptation add an additional factor of uncertainty, with different classes of uncertainty (imprecision, uncertainty, etc), each having a different effect on adaptation. E.g., when considering smart energy systems, factors could be:

Uncertainty: How much energy is injected into the power network and when
Imprecision: How much energy will windmills produce taking into account the weather forecast
Unforeseen: How the system is under threat from a cyber-attack

Obviously, when dealing with adaption and evolution, there is a strong interdependency with *models* of CPS. Specifically, the adding models (of environment or system parts) to the system to drive adaptation is a relevant issue. Here, models can be added manually, or even autonomously (e.g., models or parameters of models can be learned). Typical scenarios in the smart energy setting include the support for user profiles, the use of load-driven cost model, or the provision of an Intrusion detection system (signature based or anomaly based).

Furthermore, there is also a strong interdependency with the issue of *design-spaces (design time)* or *configuration spaces (run time)*. On the one hand the design space has an impact on possible directions of adaptations; furthermore, system evolvability also in turn impacts the design space. As construction and operation of the system blur, the design and the configuration space merge into solution space. Using the smart energy context again, it becomes obvious that if the interface between households and energy network does not describe household controlled mechanisms for offloading to the grid, the configuration space cannot be used to direct the adaptation.

Finally, there also is a strong interdependency to the issue of *composition*. Here, the classical notion is too static to deal with adaptation. E.g. when considering certification, current composition approaches require to reanalyze / re-certify the complete system after change. Specifically, current approaches to not take into account the different new levels of composition (e.g. thermal, electrical, temporal) in CPS. As a result, the classical separation of concerns/domains incl. implicit assumptions/decisions (e.g. mechanics, HW, SW) limits possible forms of adaptations.

5 Open Problems

Cyber-physical systems are engineered systems created as networks of interacting physical and computational processes. Most modern products in major industrial sectors, such as automotive, avionics, medical devices or energy production and distribution already are or rapidly become CPS driven by new requirements and competitive pressures. Science and technology advancements in the 20th century have produced methods and tools for designing computational and physical systems in isolation. However, these methods have proved to be inadequate in a large range of CPS, where computational and physical processes are so tightly integrated that it is not possible to identify whether behavioral characteristics are the result of computations (computer programs), physical laws, or both working together, and where functionality and salient system characteristics are emerging through the interaction of physical and computational objects. CPS research targets the establishment of a new

system science that reintegrates physical and information sciences and creates new science and technology foundations for CPS that is simultaneously physical and computational. There are many open problems whose solutions will guide progress toward this new systems science. Examples for these open problems are the following:

1. **New abstractions for CPS design flows.** Heterogeneity is the norm as well as the main challenge in CPS design: components and systems are modeled using multiple physical, logical, functional and non-functional modeling aspects. The scope of relevant design domains includes (1) multiple physical domains, such as 3D structure, mechanical, thermal, fluid, electrical, electromagnetic and (2) computational/networking domains, such as system control, sensors, health management, mission management, communication. Modeling and analyzing cross-domain interactions among physical and computational/networking domains and understanding the effects of heterogeneous abstraction layers in the design flow are fundamental part of CPS design theories.
2. **Semantic foundations for composing heterogeneous models and modeling languages** describing different physics and logics. Design automation for CPS requires the introduction of mathematical frameworks that make semantics not only mathematically precise, but also explicit, understandable and practical for system developers as well as tool developers.
3. **Compositionality in heterogeneous systems** that allows taking into account both physical and computational properties is an open problem. This new view of compositionality is required to create large, networked systems that satisfy essential physical properties and deliver the required functionality in a reliable way.
4. Cyber physical systems will have properties for which achieving full compositionality would be expensive or impractical. Development of technology for achieving **predictability in partially compositional properties** is a hard problem that must be addressed.
5. **Scientific foundation for system integration** that is model-based, precise, and predictable. Transforming system integration from a high risk engineering practice into a science-based engineering discipline is a huge challenge that will require close collaboration between industry and academy.
6. **Compositional certification of cyber-physical systems.** New theories and methods are needed for composing CPS components into a large CPS system in such a way that the certification of the components can be reused as evidence in certifying the larger system.
7. **Agile design automation** of cyber-physical systems. As new CPS application domains appear, the existing tool base needs to be rapidly adapted to the new systems. If companies must wait for tools to be created before we they can move into new areas, they will lose the lead to competitors who can use either agile tool chains or massive amounts of labor to work through design problems.
8. **Resilient CPS systems** that can tolerate malicious attacks from either the cyber or physical domains. New architectures, model-based design methods and tools are required to build resilient systems.

Solution to these open problems will enable new generations of CPS products and rapid progress in the technology infrastructure of modern societies.

Participants

- Luis Almeida
University of Porto, PT
- Panos J. Antsaklis
University of Notre Dame, US
- Karl-Erik Arzén
Lund University, SE
- Gary Balas
University of Minnesota, US
- Basil Becker
Hasso-Plattner-Institut – Potsdam, DE
- Kirstie Bellman
Aerospace Corp. – Los Angeles, US
- Frederic Boulanger
Supélec – Gif-sur-Yvette, FR
- Christian Buckl
fortiss GmbH – München, DE
- John S. Fitzgerald
Newcastle University, GB
- Martin Fränzle
Universität Oldenburg, DE
- Holger Giese
Hasso-Plattner-Institut – Potsdam, DE
- Cecile Hardebolle
Supélec – Gif-sur-Yvette, FR
- Constance L. Heitmeyer
Naval Res. – Washington, US
- Andreas Herkersdorf
TU München, DE
- Jérôme Hugues
ISAE – Toulouse, FR
- Michaela Huhn
TU Clausthal, DE
- Hardi Hungar
OFFIS – Oldenburg, DE
- Gabor Karsai
Vanderbilt University, US
- Philip Koopman
Carnegie Mellon University – Pittsburgh, US
- Stefan Kowalewski
RWTH Aachen, DE
- Christoph Krauß
Fraunhofer AISEC – München, DE
- Peter Bernard Ladkin
Universität Bielefeld, DE
- Tom S. Maibaum
McMaster Univ. – Hamilton, CA
- Pieter J. Mosterman
The MathWorks Inc. – Natick, US
- Jerzy W. Rozenblit
Univ. of Arizona – Tucson, US
- Bernhard Rumpe
RWTH Aachen, DE
- Bernhard Schätz
fortiss GmbH – München, DE
- Janos Sztipanovits
Vanderbilt University, US
- Hans Vangheluwe
McGill University, CA, and University of Antwerp, BE
- Alan Wassyng
McMaster Univ. – Hamilton, CA

Report from Dagstuhl Seminar 11451

Data Mining, Networks and Dynamics

Edited by

Lars Eldén[1] and Andreas Frommer[2]

1 Linköping University, SE, `lars.elden@liu.se`
2 Universität Wuppertal, DE, `Frommer@math.uni-wuppertal.de`

Abstract

In many areas one needs to extract relevant information from signals generated by dynamical systems evolving on networks with a configuration that itself evolves with time. Such problems occur e.g. in surveillance systems for security, in early warning systems of disasters such as earthquakes, hurricanes and forest fires, in collaborative filtering, in search engines for evolving data bases and in reputation systems based on evolving votes. In each of those examples one wants to extract relevant information from an evolving database. Information science is one of the most expansive scientific areas nowadays, much due to the vast amount of information that is available on the Internet, and and the rapid growth of e-business. There are several open problems in these areas of research and even partial answers would have an important impact. There is a pressing need to make progress in analysis tools and in algorithms for such complex tasks. The purpose of the Dagstuhl Seminar 11451 "Data Mining, Networks and Dynamics" was to bring together a diverse community of researchers working in different aspects of this exciting field.

Seminar 06.–11. November, 2011 – www.dagstuhl.de/11451
1998 ACM Subject Classification G 1.3 Numerical linear algebra, G 1.6 Optimization, H 3.3 Information search and retrieval
Keywords and phrases dynamic networks, data mining, optimization on manifolds, iterative methods, graph algorithms, network models in applications
Digital Object Identifier 10.4230/DagRep.1.11.23
Edited in cooperation with Karsten Kahl

1 Executive Summary

Lars Eldén
Andreas Frommer

The main focus of the Dagstuhl Seminar 11451 "Data Mining, Networks and Dynamics" was the theory and the computational aspects of methods for the extraction of information in evolving networks and recent advances in algorithms for related linear algebra problems. The ever increasing size of data sets and its influence on algorithmic progress appeared as a recurrent general theme of the seminar. Some of the participants are experts in the modeling aspects, some are focusing on the theoretical analysis, and some are more directed toward software development and concrete applications. There was a healthy mix of participants of different age and academic status, from several PhD students and post-docs to senior researchers. As the subject has immediate and important applications, the seminar had some attendants with an industrial background (Yahoo and Google). Those participants

Data Mining, Networks and Dynamics, *Dagstuhl Reports*, Vol. 1, Issue 11, pp. 23–38
Editors: Lars Eldén and Andreas Frommer
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

also contributed greatly by introducing the academic researchers to new applications. The seminar also had several academic participants from application areas, who presented recent advances and new problems. Apart from well-known applications in social networks, search engines and biology approached from a different angle (Alter, Bast, Groh, Harb, Stumme), new applications were presented such as network methods in epidemiology (Poletto), human contact networks (Yoneki), credibility analysis of Twitter postings (Castillo), structure determination in cryo-electron microscopy (Boumal).

On the methodological side, new methods from graph theory and corresponding numerical linear algebra were presented (Bolten, Brannick, Delvenne, Dhillon, Gleich, Ishteva, Kahl, Savas, Szyld). Of particular interest is the extraction of information from extremely large graphs. Given that the class of multigrid methods is standard for solving large sparse matrix problems derived from partial differential equations, it is very natural that these methods should be tried for graph problems. In this direction, new adaptive algebraic multigrid methods for obtaining the stationary distribution of Markov chains were presented.

It has recently been recognized that optimization on manifolds (Boumal, Sepulchre) is a powerful tool for solving problems that occur in information sciences. As real world data are often organized in more than two categories, tensor methods (Alter, De Lathauwer, Eldén, Khoromskij, Sorber) are becoming a hot topic, and the talks showed that the techniques are developing so that now large problems can be treated. Tensor methods have been used for a long time for extremely large problems arising in physics. It has been conjectured that those techniques can be used also for problems in information science. Preliminary discussions along those lines took place. It is also interesting that some tensor computations are based on manifold optimization. Thus there were interesting discussions on the interplay between these areas.

The atmosphere of the meeting was very informal and friendly. During and after the talks lively discussions took place that also continued after dinner. Although it is too early to tell whether the seminar lead to new collaborations between the participants, some preliminary contacts were made. An open problem in spectral partitioning was raised (Eldén) and a preliminary solution was suggested (Gleich, Kahl).

The participants of this seminar had a chance to interact with the Dagstuhl seminar 11542 "Analysis of Dynamic Social and Technological Networks" held at the same time. Indeed, the Thursday morning session was arranged as a common session between both seminars, focusing at introducing the different methodological approaches to all particpants.

2 Table of Contents

3 Overview of Talks

3.1 Discovery of Mechanisms and Prognosis of Cancers from Matrix and Tensor Modeling of Large-Scale Molecular Biological Data

Orly Alter (University of Utah – Salt Lake City, US)

In my Genomic Signal Processing Lab, we believe that future discovery and control in biology and medicine will come from the mathematical modeling of large-scale molecular biological data, such as DNA microarray data, just as Kepler discovered the laws of planetary motion by using mathematics to describe trends in astronomical data [1]. In this talk, I will first describe novel generalizations of the matrix and tensor computations that underlie theoretical physics (e.g., [2, 3]), that we are developing for comparison and integration of multiple genomic datasets recording different aspects of, e.g., cell division and cancer. Second, I will describe our recent experiments [4] that verify a computationally predicted genome-wide mode of regulation [5, 6], and demonstrate that singular value decomposition (SVD) and higher-order SVD (HOSVD) modeling of DNA microarray data can be used to correctly predict previously unknown cellular mechanisms. Third, I will show that mode-1 HOSVD modeling of rRNA sequence alignments suggests a new way of looking at evolution as a composition of changes rather than a hierarchy of changes, and might be used to predict evolutionary mechanisms, i.e., evolutionary pathways and the underlying structural changes that these pathways are correlated, possibly even coordinated with [7]. Last, I will describe the computational prognosis of brain cancers by using generalized SVD to compare global DNA copy numbers in patient-matched normal and tumor samples from the Cancer Genome Atlas [8, 9].

References

1 O. Alter, PNAS 103, 16063–16064 (2006); http://dx.doi.org/10.1073/pnas.0607650103
2 O. Alter, P. Brown, and D. Botstein, PNAS 100, 3351–3356 (2003); http://dx.doi.org/10.1073/pnas.0530258100
3 S. P. Ponnapalli, M. Saunders, C. Van Loan, and O. Alter, PLoS One 6 (12): e28072 (2011); http://dx.doi.org/10.1371/journal.pone.0028072
4 L. Omberg, J. Meyerson, K. Kobayashi, L. Drury, J. Diffley, and O. Alter, MSB 5, 312 (2009); http://dx.doi.org/10.1038/msb.2009.70
5 O. Alter, and G. Golub, PNAS 101, 16577–16582 (2004); http://dx.doi.org/10.1073/pnas.0406767101
6 L. Omberg, G. Golub, and O. Alter, PNAS 104, 18371–18376 (2007); http://dx.doi.org/10.1073/pnas.0709146104
7 C. Muralidhara, A. Gross, R. Gutell, and O. Alter, PLoS One 6 (4), e18768 (2011); http://dx.doi.org/10.1371/journal.pone.0018768
8 C. Lee, and O. Alter, 60th Annual Meeting of the American Society of Human Genetics (Washington DC, November 2–6, 2010).
9 C. Lee, B. Alpert, P. Sankaranarayanan, and O. Alter, PLoS One 7 (1), e30098 (2012); http://dx.doi.org/10.1371/journal.pone.0030098

3.2 Semantic Full-Text Search

Hannah Bast (Universität Freiburg, DE)

Joint work of Bast, Hannah; Bäurle, Florian; Buchhold, Björn; Haußmann, Elmar

Semantic full-text search combines classical full-text search with search in knowledge databases aka ontologies. An example query would be to find all female german professors working in a particular research area. Google will not be of much help for such a query, unless there is a web page with exactly that contents. And a general-purpose ontology is unlikely to contain all the information required to answer this query.

I will explain this in more detail and how it differs from other approaches claiming to be "semantic". I will point out some of the underlying algorithmic problems, especially with respect to scalability, and some of our current solutions. And I will demonstrate our current prototype. All this is fairly recent and, as I like to think, exciting work in progress.

3.3 Multigrid methods for circulant and Toeplitz matrices and their application to queueing networks

Matthias Bolten (Universität Wuppertal, DE)

Toeplitz and circulant preconditioners are used to speed up Krylov subspace methods in queueing applications. While systems involving circulant matrices can be solved in near-optimal, i.e. $O(N \log N)$, time, this is not the case for Toeplitz matrices. Nevertheless, circulant preconditioners often do not provide the required performance. In this talk we will introduce multigrid methods for Toeplitz and circulant matrices that overcome some of these limitations. The convergence theory, including results for aggregation and smoothed aggregation, will be presented and their application to queueing networks will be discussed.

3.4 Synchronization of rotations on a graph

Nicolas Boumal (UC Louvain-la-Neuve, BE)

Joint work of Boumal, Nicolas; Singer, Amit; Absil, Pierre-Antoine
Main reference A. Singer, "Angular synchronization by eigenvectors and semidefinite programming, Applied and Computational Harmonic Analysis," Vol. 30, Issue 1, January 2011, pp. 20–36.
URL http://dx.doi.org/10.1016/j.acha.2010.02.001

I will start by informally describing two application problems. These will motivate the study of an estimation problem in which the quantities to estimate are N rotations and the measurements to help us do so are noisy observations of relative rotations. The measurements induce a graph which consists of N nodes (each associated to one of the unknown rotations) and such that each edge corresponds to a noisy measurement of the relative rotation between the two incident nodes. I will then state a few results we have obtained on the theoretical

and algorithmic aspects of this problem, and try to gradually transform the presentation into a group discussion.

3.5 Algebraic Distance based Coarsening with Application to Clustering

James Brannick (Pennsylvania State University, US)

In this talk I introduce an algebraic distance technique for multilevel graph organization. The ideas build on recent work concerning development of algebraic multigrid solvers for linear systems, in particular elliptic partial differential equations.

3.6 Information Credibility on Twitter

Carlos Castillo (Yahoo Research – Barcelona, ES)

Joint work of Castillo, Carlos; Mendoza, Marcelo; Poblete, Bárbara
Main reference C. Castillo, M. Mendoza, B. Poblete, "Information credibility on twitter," in Proc. of the 20th Int'l Conf. on World Wide Web (WWW'11), pp. 675–684, ACM, New York, NY, USA, 2011.
URL http://dx.doi.org/10.1145/1963405.1963500

We analyze the information credibility of news propagated through Twitter, a popular microblogging service. Previous research has shown that most of the messages posted on Twitter are truthful, but the service is also used to spread misinformation and false rumors, often unintentionally.

On this paper we focus on automatic methods for assessing the credibility of a given set of tweets. Specifically, we analyze microblog postings related to "trending" topics, and classify them as credible or not credible, based on features extracted from them. We use features from users' posting and re- posting ("re-tweeting") behavior, from the text of the posts, and from citations to external sources.

We evaluate our methods using a significant number of human assessments about the credibility of items on a recent sample of Twitter postings. Our results shows that there are measurable differences in the way messages propagate, that can be used to classify them automatically as credible or not credible, with precision and recall in the range of 70% to 80%.

3.7 Block Term Decompositions and Block Component Analysis

Lieven De Lathauwer (K.U. Leuven, BE)

The Canonical Polyadic Decomposition (CPD) of a higher-order tensor is rank-revealing; it writes the tensor as a minimal sum of rank-1 terms. The Tucker decomposition finds the column space, row space, ... of a higher-order tensor and is multilinear rank-revealing.

The recently introduced Block Term Decompositions (BTD) generalize these two basic decompositions. They write a higher-order tensor as a sum of terms that have low multilinear rank. BTDs enable a new approach to factor analysis and blind source separation, which we call Block Component Analysis. We discuss links with Canonical Polyadic Analysis (CPA), Independent Component Analysis (ICA), Pareto analysis and compressed sensing. We show several illustrative examples.

3.8 Dynamics on networks for communities and centralities

Jean-Charles Delvenne (UC Louvain, BE)

Joint work of Delvenne, Jean-Charles; Yaliraki, Sophia; Barahona, Mauricio
Main reference J.-C. Delvenne, S. Yaliraki, M. Barahona, "Stability of graph communities across time scales," in Proc. of the National Academy of Sciences, 107:29, pp. 12755–12760.
URL http://dx.doi.org/10.1073/pnas.0903215107

Dynamical systems taking place on networks, such as opinion dynamics, synchronisation, consensus or random walks, reveal a lot about their structure.

In particular we show, through a dynamical reinterpretation of well-known concepts, how centrality measures ('which nodes are most central in the network?') and community detection quality functions ('which group of nodes are tightly related to one another?') are intimately related. The dynamical interpretation allows to design new centrality or community quality measures tailored for every particular application.

This is joint work with Sophia Yaliraki, Mauricio Barahona, Renaud Lambiotte, Anne-Sophie Libert [1, 3, 2].

References

1 J.-Ch. Delvenne, S.N. Yaliraki, and M. Barahona. Stability of graph communities across time scales. *Proc. Natl Acad. Sci. U. S. A.*, 107:12755–12760, 2010.
2 J.C. Delvenne and A.S. Libert. Centrality measures and thermodynamic formalism for complex networks. *Physical Review E*, 83(4):046117, 2011.
3 R. Lambiotte, J.-C. Delvenne, and M. Barahona. Dynamics and modular structure in networks. *arXiv:0812.1770*, 2008.

3.9 Fast and memory-efficient low rank approximation of massive graphs

Inderjit Dhillon (University of Texas – Austin, US)

Joint work of Dhillon, Inderjit; Savas, Berkant

Social network analysis requires us to perform a variety of analysis tasks, including summarization and prediction, on massive graphs. In this talk, I will present a fast and memory-efficient procedure called clustered low rank matrix approximation for massive graphs. The procedure involves a fast clustering of the graph followed by approximation of each cluster separately using existing methods, e.g. the singular value decomposition, or stochastic algorithms. The clusterwise approximations are then extended to approximate the entire graph.

This approach has several benefits: (1) important structure of the graph is preserved due to the clustering; (2) accurate low rank approximations are achieved; (3) the procedure is efficient both in terms of computational speed and memory usage. Further, we generalize stochastic algorithms into the clustered low rank approximation framework and present theoretical bounds for the approximation error. Finally, a set of experiments show that our methods outperform existing dimensionality reduction algorithms on massive social networks.

3.10 Simultaneous clustering of time-evolving graphs

Lars Eldén (Linköping University, SE)

Spectral partitioning is a well-known method for the clustering and partitioning of graphs. The Fiedler vector of the graph Laplacian is used to reorder the vertices of the graph. This makes it possible to find a partitioning that is rather close to being optimal. Alternatively the Fiedler vector can be computed using the adjacency matrix of the graph. We discuss the generalization of this method to tensors, which made up from a graph that evolves with time. Our approach is based on a generalization of the Rayleigh quotient characterization of the Fiedler vector. The solution of the tensor Rayleigh quotient maximization problem is computed using a tensor-Krylov method. A couple of examples are given, with a clustering of a sparse tensor obtained from the Enron email data set.

3.11 Techniques for local and global centrality estimation

David F. Gleich (Purdue University, US)

Joint work of Bonchi, Francesco; Esfandiar, Pooya; Gleich, David F.; Greif, Chen; Lakshmanan, Laks V. S.
Main reference F. Bonchi, P. Esfandiar, D.F. Gleich, C. Greif, L.V.S. Lakshmanan, "Fast matrix computations for pair-wise and column-wise Katz scores and commute times," arXiv:1104.3791v1 [cs.SI]; to appear in J. Internet Mathematics.
URL http://arxiv.org/abs/1104.3791v1

In this talk, I will discuss local algorithms for centrality scores, with a particular focus on the Katz score. I will also discuss recent work on how to utilize overlapping partitions of a graph to reduce the communication in the computation of a PageRank or Katz vector. Both of these discussions will focus on static graphs. In the last third of the talk, I present thoughts on adapting these algorithms to time-dependent problems and show some curious properties of graph spectra I cannot explain.

3.12 Social Situation Detection on the Basis of Interaction Geometries and Contextual Privacy Management as an Application

Georg Groh (TU München, DE)

Joint work of Groh, Georg; Fuchs, Christoph; Lehmann, Alexander; Birnkammerer, Stefan

Social Situation models are special forms of episodic short term social context that can be detected with the help of mobile sensors and used for applications in a (distributed, mobile) Social Networking scenario. The talk centers around methods to detect Social Situations by analyzing geometry of interaction (e.g. shoulder angles, interpersonal distances), audio or other sensor data sources. A special approach for privacy management is introduced as an application using spatio-temporal contexts as well as short term and long term social contexts as building blocks.

References

1 Groh, G., Fuchs, C., and Lehmann, A. (2011d). Combining evidence for social situation detection. *Proc. IEEE Socialcom'11, Boston, USA*.
2 Groh, G. and Lehmann, A. (2011). Deducing evidence for social situations from dynamic geometric interaction data. *Int. J. Social Computing and Cyber-Physical Systems (to appear)*.
3 Groh, G. and Birnkammerer, S. (2011). Privacy and information markets: Controlling information flows in decentralized social networking. *Proc. IEEE Socialcom'11, Boston, USA*.

3.13 Extracting, Searching and Discovering the Semantics of Tables on the Web

Boulos Harb (Google – New York, US)

Joint work of Das Sarma, Anish; Halevy, Alon; Harb, Boulos; Lee, Hongrae; Wu, Fei; Yu, Cong
Main reference P. Venetis, A. Halevy, J. Madhavan, M. Pasca, W. Shen, F. Wu, G. Miao, C. Wu, "Recovering semantics of tables on the web," in Proc. VLDB Endow., 4 (2011), pp. 528–538.
URL http://dl.acm.org/citation.cfm?id=2002939

The Web provides a large corpus of over 100 million high-quality tables that offer a rich source of structured information on an extensive variety of topics. However, the semantics of each of these HTML tables are rarely explicit in the table itself, and the meaning is often derived from text surrounding the table. Header rows exist in only a few cases and even when they do, the attribute names are not standardized and typically of not much use. Without understanding the attributes and entities a table represents it is difficult to better leverage its content (in search for example) or join or union it with other tables. We have built a scalable system for automatically recovering the semantics of tables on the Web. Given the scale, breadth and heterogeneity of this corpus, we cannot rely on hand-coded domain knowledge to associate semantics with each table. To recover semantics of tables, we leverage a (noisy) database of class labels and relationships automatically extracted from the Web. We add annotations to a table describing the sets of entities and attributes (represented in the rows and columns respectively) the table is modeling, and the binary relationships between the attribute and the subject columns. Our system is extensible to

any corpus of tables, and we leverage it to build a search over public Google Fusion Tables (http://www.google.com/fusiontables).

3.14 Low rank nonnegative approximations

Mariya Ishteva (Georgia Institute of Technology, US)

Joint work of Ishteva, Mariya; De Lathauwer, Lieven; Park, Haesun

Nonnegative matrix factorization (NMF) is an (approximate) factorization of a nonnegative matrix into two nonnegative matrices with low rank. It has proven to be a successful dimensionality reduction and factor analysis technique.

We consider a related problem where we do not impose nonnegativity on each factor but only on their product. This can be thought of as a low rank nonnegative approximation of a given matrix. In many applications, e.g., image compression, the NMF constraints are too strong if we only need a nonnegativity in the product. Our approximation combines advantages of both NMF and singular value decomposition (SVD). The approximation error is theoretically not larger than that of NMF (and not smaller than that of SVD) and in many experiments we observed that it is much closer to that of SVD especially when the reduced dimension is small. On the other hand, because of the nonnegativity constraints, the approximation has potentially better interpretability than the one obtained from SVD.

We propose an active-set based algorithm for the computation of the factorization using the Frobenius norm as a distance measure. We also show how the problem of approximating a higher-order nonnegative tensor by a low multilinear rank nonnegative tensor can be reduced to its matrix counterpart.

3.15 Bootstrap Algebraic Multigrid for Markov Chains

Karsten Kahl (Universität Wuppertal, DE)

This talk is devoted to the development of an adaptive algebraic multilevel method for computing steady state vectors of Markov chains. We present an efficient bootstrap algebraic multigrid method for this task.

In our proposed approach, we employ a multilevel eigensolver, with interpolation built using ideas based on compatible relaxation, algebraic distances, and least squares fitting of test vectors. Our adaptive variational strategy for computation of the steady state vector of a given Markov chain is then a combination of this multilevel eigensolver and an associated additive multilevel preconditioned correction process. We show that the bootstrap algebraic multigrid eigensolver by itself can efficiently compute accurate approximations to the steady state vector. An additional benefit of the bootstrap approach is that it yields an accurate interpolation operator for many other eigenmodes. This in turn allows for the use of the resulting multigrid hierarchy as a preconditioner to accelerate the GMRES iteration for computing an additive correction equation for the approximation to the steady state vector. Unlike other existing multilevel methods for Markov chains, our method does not employ

any special processing of the coarse-level systems to ensure that stochastic properties of the fine-level system are maintained there. The proposed approach is applied to a range of test problems, involving non-symmetric M-matrices arising from stochastic matrices, showing promising results.

3.16 Tensor numerical methods for multi-dimensional PDEs: application to parabolic equations

Boris N. Khoromskij (MPI für Mathematik in den Naturwissenschaften, DE)

Modern methods of rank-structured tensor decomposition allow an efficient data-sparse separable approximation of multivariate functions and operators, providing linear complexity scaling in the dimension.

The successful applications of tensor methods include the challenging high-dimensional problems arising in material sciences, bio-science, stochastic modeling, signal processing, machine learning and data mining, quantum computing, many-body dynamics etc.

The recent quantized tensor train (QTT) technique is proved to provide the super-compressed representation for a class of multidimensional data-arrays of size N^d with the log-volume storage complexity $O(d \log N)$.

The approach is based on data quantization to the $2 \times 2 \times ... \times 2$-format in the highest possible dimension $D = d \log N$.

Combined with the efficient multi-linear QTT-algebra, this method opens the way to the profound deterministic numerical simulation of high-dimensional PDEs getting rid of the "curse of dimensionality" and restrictions on the grid size N.

However, in some cases, the approach is limited by the "curse of ranks" characterizing the essential entanglement in the system of interest.

We discuss the asymptotically optimal low QTT-rank representations for a class of multivariate functions and operators (matrices), focusing on multivariate potentials, d-dimensional elliptic Green's function and parabolic solution operator.

Numerical illustrations in electronic structure calculations (the Hartree-Fock equation), quantum molecular dynamics (the molecular Schrödinger equation) and stochastic modeling for dynamics of liquids (the Fokker-Planck equation), are presented.

3.17 Multiscale Networks and the Spatial Spread of Infectious Diseases

Chiara Poletto (ISI Foundation, IT)

Human mobility and interactions represent key ingredients in the spreading dynamics of an infectious disease. The flows of traveling individuals form a network characterized by complex features, such as strong topological and traffic heterogeneities, that unfolds at different temporal and spatial scales, from short ranges to the global scale. Computational models can be developed that integrate detailed network structures based on demographic and mobility data, in order to simulate the spatial evolution of an epidemic. Focusing on the 2009 H1N1

influenza pandemic as a paradigmatic example, these approaches allow quantifying the effects of travel restrictions in delaying and controlling the epidemic spread. In addition, simplified model frameworks can be solved providing the assessment of the interplay between individual mobility and epidemic dynamics, under different conditions of heterogeneities characterizing human mobility.

3.18 Supervised Link Prediction Using Multiple Sources

Berkant Savas (Linköping University, SE)

Joint work of Savas, Berkant; Lu, Zhengdong; Tang, Wei; Dhillon, Inderjit S.
Main reference Z. Lu, B. Savas, W. Tang, I. S. Dhillon, "Link prediction using multiple sources of information," in Proc. of the IEEE Int'l Conf. on Data Mining (ICDM'10), pp. 923–928, 2010.
URL http://dx.doi.org/10.1109/ICDM.2010.112

Link prediction is a fundamental problem in social network analysis and modern-day commercial applications. Most existing research approaches this problem by exploring the topological structure of a social network using only one source of information. However, in many application domains, in addition to the social network of interest, there are a number of auxiliary social networks and/or derived proximity networks available. In this talk we will present: (1) a supervised learning framework that can effectively and efficiently learn the dynamics of social networks in the presence of auxiliary networks; (2) a feature design scheme for constructing a rich variety of path-based features using multiple sources, and an effective feature selection strategy based on structured sparsity. Extensive experiments on three real-world collaboration networks show that our model can effectively learn to predict new links using multiple sources, yielding higher prediction accuracy than unsupervised and single source supervised models.

3.19 Computing with fixed (low) rank matrices: a geometric approach

Rodolphe Sepulchre (University of Liège, BE)

Joint work of Sepulchre, Rodolphe;Bonnabel, Silvère; Meyer, Gilles; Mishra, Bamdev; Absil, PA; Vandereycken, B.
Main reference G. Meyer, S. Bonnabel, R. Sepulchre, "Regression on fixed-rank positive semidefinite matrices: a Riemannian approach," Journal of Machine Learning Research, 12(Feb):593–625, 2011.
URL http://www.montefiore.ulg.ac.be/~meyer/papers.php

The talk is an introduction to a recent computational framework for optimization over the set of fixed rank positive semidefinite matrices. The proposed framework is based on quotient Riemannian geometries that lead to rank-preserving efficient computations in the set of fixed-rank matrices. The field of applications is vast, and the talk will survey recent developments that illustrate the potential of the approach in large-scale computational problems encountered in control, optimization, and machine learning.

3.20 Optimization-based algorithms for the decomposition of a tensor in rank-(Lt,Lt,1) terms

Laurent Sorber (K.U. Leuven, BE)

Joint work of Sorber, Laurent; Van Barel, Marc; De Lathauwer, Lieven

The canonical polyadic and rank-(Lt,Lt,1) block term decomposition (CPD and BTD, respectively) are two closely related tensor decompositions. The CPD is an important tool in psychometrics, chemometrics, neuroscience and data mining, while the rank-(Lt,Lt,1) BTD is an emerging decomposition in signal processing and, recently, blind source separation. We present a decomposition that generalizes these two and develop algorithms for its computation. Among these algorithms are alternating least squares schemes, several general unconstrained optimization techniques, as well as matrix-free nonlinear least squares methods. In the latter we exploit the structure of the Jacobian's Gramian by means of efficient expressions for its matrix-vector product. Combined with an effective preconditioner, numerical experiments confirm that these methods are among the most efficient and robust currently available.

3.21 Data Mining in Online and Offline Social Networks

Gerd Stumme (Universität Kassel, DE)

This talk presents selected results of the analysis of two real-world applications: BibSonomy and Conferator.

BibSonomy (http://www.bibsonomy.org) is a social bookmark and publication sharing system that is run by the Knowledge & Data Engineering Group of the University of Kassel. We will discuss how to obtain a conceptual hierarchical clustering of the folksonomy of the system by means of Formal Concept Analysis, and how an adapted version of PageRank, called FolkRank, will improve the search within the system and allow for the analyis of time series. We will also briefly touch the topics of ontology learning and spam detection.

Conferator (http://conferator.org) gives you the opportunity to personalize the conference schedule, and shows you a history of your social contacts at the conference and provides additional information about participants, such as their homepage, facebook and twitter accounts, contact details (skype, icq, etc.).

After the conference, Conferator will thus enable you to recall your social contacts you had during the conference. We will discuss a variety of approaches for detecting the most central participants of the conference.

3.22 Matrices with Perron-Frobenius Properties

Daniel B. Szyld (Temple University – Philadelphia, US)

Joint work of Szyld, Daniel B.; Elhashash, Abed
Main reference A. Elhashash, D.B. Szyld, "Two characterizations of matrices with the Perron-Frobenius property," Numerical Linear Algebra with Applications, vol. 16, pp. 863–869, 2009.
URL http://www.math.temple.edu/szyld

The Perron-Frobenius theorem guarantees the existence of a nonnegative (or positive) eigenvector corresponding to the dominant eigenvalue of a nonnegative matrix. In this talk, we explore the question: which matrices (not necessarily nonnegative) also have such a vector? We provide combinatorial, algebraic and spectral characterizations of such matrices. We also generalize the concept of M-matrix using these Perron Frobenius properties of matrices.

3.23 Wireless Epidemic in Dynamic Human Contact Networks

Eiko Yoneki (University of Cambridge, GB)

Joint work of Yoneki, Eiko; Hui, P.; Crowcroft, J.
Main reference P. Hui, J. Crowcroft, E. Yoneki, "BUBBLE Rap: Social-based Forwarding in Delay Tolerant Networks," IEEE Transactions on Mobile Computing, vol. 10, no. 11, pp. 1576–1589, 2011.
URL http://dx.doi.org/10.1109/TMC.2010.246

Increasing numbers of mobile computing devices form dynamic time-dependent networks in daily life. We have explored new communication paradigms using dynamic interconnectedness between devices. To reduce the overhead of epidemic routing in such networks, we attempt to uncover a hidden stable network structure such as a social network, consisting of a group of people forming meaningful relationships (community) and centrality nodes, that act as rely nodes for message passing. We have exploited device connectivity traces from the real world for modelling social network structure. By using these traces, we have shown that social network properties are an important aspect of building a routing protocol.

The empirical study of contact networks shares many issues with network-based epidemiology, and our work has been extended towards understanding the epidemic spread of infectious diseases. Capturing human interactions will provide an empirical, quantitative measurement of social mixing patterns to underpin mathematical models of the spread of close-contact diseases. We have extracted multi-spread modes of operation from contact networks.

Participants

- Orly Alter
University of Utah –
Salt Lake City, US
- Hannah Bast
Universität Freiburg, DE
- Matthias Bolten
Universität Wuppertal, DE
- Nicolas Boumal
UC Louvain-la-Neuve, BE
- James Brannick
Pennsylvania State Univ. US
- Carlos Castillo
Yahoo Research – Barcelona, ES
- Lieven De Lathauwer
K.U. Leuven, BE
- Jean-Charles Delvenne
UC Louvain, BE
- Inderjit Dhillon
University of Texas – Austin, US
- Lars Eldén
Linköping University, SE
- Andreas Frommer
Universität Wuppertal, DE
- David F. Gleich
Purdue University, US
- Georg Groh
TU München, DE
- Boulos Harb
Google – New York, US
- Mariya Ishteva
Georgia Inst. of Technology, US
- Karsten Kahl
Universität Wuppertal, DE
- Boris N. Khoromskij
MPI für Mathematik in den
Naturwissenschaften, DE
- Chiara Poletto
ISI Foundation, IT
- Berkant Savas
Linköping University, SE
- Rodolphe Sepulchre
University of Liège, BE
- Alexander J. Smola
Yahoo! Res. – Santa Clara, US
- Laurent Sorber
K.U. Leuven, BE
- Gerd Stumme
Universität Kassel, DE
- Daniel B. Szyld
Temple Univ. – Philadelphia, US
- Eiko Yoneki
University of Cambridge, GB

Report from Dagstuhl Seminar 11452

Analysis of Dynamic Social and Technological Networks

Edited by

Vito Latora[1], Cecilia Mascolo[2], and Mirco Musolesi[3]

1 **Università di Catania, IT,** latora@ct.infn.it
2 **University of Cambridge, GB,** cecilia.mascolo@cl.cam.ac.uk
3 **University of Birmingham, GB,** m.musolesi@cs.bham.ac.uk

Abstract

With the growing popularity of online communication tools, researchers have turned their attention to the study of the networks arising between users of social networking services, between mobile phone callers and, in general, between individuals connected by technological means. Thanks to the rich set of techniques and methods developed by complex network science, and joining forces with sociologists and psychologists, the analysis of dynamic social and technological networks has sparked many important results, attracting even more interest as the importance of such systems grows over time.

This Dagstuhl seminar brought together researchers and practitioners from computer science, physics and psychology, covering the diverse areas of social and technological network analysis. The goal of the seminar was to bring together people from different areas of expertise, focusing on both mathematical aspects and practical applications of theoretical models and techniques. In particular, the evolution of this research field and of its future perspectives was a major theme of the seminar. This seminar was attended by 25 participants.

Seminar 8.–11. November, 2011 – www.dagstuhl.de/11452
1998 ACM Subject Classification H.2.8 Data Mining.
Keywords and phrases complex networks, network analysis, network data mining
Digital Object Identifier 10.4230/DagRep.1.11.39
Edited in cooperation with Salvatore Scellato

1 Executive Summary

Vito Latora
Cecilia Mascolo
Mirco Musolesi

In the recent years, we have witnessed an increasing interest in the analysis of complex networks, i.e., networks composed of many interacting entities that show emergent behavior at global level. Usually, the key features of these networks can be captured by means of a statistical characterisation of their properties at local level such as their degree distribution and clustering coefficient. Models can then be built in order to describe the system in its entirety and to study the processes taking place on it, also over time. These models can be used to understand several phenomena dependent on the structure and dynamics of these networked systems, such as the spreading of computer viruses.

One of the main motivations of the explosion of interest in social networks that we are seeing is the availability of large data sets, e.g., the "snapshots" of the Internet structure or

the maps of online social networks obtained by crawling extremely popular Web sites like Facebook and Twitter. Moreover, many data collections exercises of data sets related to human interactions by means of Bluetooth radio or GPS receivers have been carried in the recent years. These large data sets provide information that is not limited to a particular instant of time, but cover a very large time interval as well as fine grained space information. These data sets can be used to study the evolution of the network and dynamic processes happening over time such as simulated epidemics. Other large data sets that have attracted considerable interest include biological, commodity and economic networks. The initial research efforts have been focussed on the analysis of the static properties of these networks, including the presence of clusters, hubs, and community structure, More recently, researchers became interested in studying dynamic processes taking place on these networks such as information diffusion in the Internet, disease epidemics and malware propagation.

The goal of this seminar was broad, including both mathematical aspects and practical applications of theoretical models and techniques. The seminar focused on two key classes of networks that are of fundamental importance not only in computer science but also in the everyday life of millions of people, namely technological networks and online social networks.

The main contributions of this seminar were:

- Presenting a wide range of recent research results on the dynamics of processes and structure of technological and social networks.
- Exchanging solutions and practices in the different areas of computer science and other disciplines in order to find novel solutions and start fruitful long-term collaborations among the seminar attendants.
- Exploring the new challenges and opportunities arising from the analysis of data from mobile devices and social network tools, which offer the change to collect very rich data sets of information about the everyday life of people including their movements, their contacts and their social network.
- Discussing how the social networks extracted from mobile device interactions are time and location dependent, requiring new models and techniques to study them.
- Examining the application of machine learning and data mining techniques to the analysis of technological and social networks, bringing together researchers and practitioners working on the massive available networking data sets and machine learning experts interested in real-world problems.
- Studying and considering the computational challenges presented by the scale of these data sets, which impose the design of novel algorithms and the rethinking of existing techniques.
- Discussing the ethical problems arising from the treatment of privacy-sensitive user data and potential technical and legal solutions to overcome them.

2 Table of Contents

3 Overview of Talks

3.1 Entropy of dynamical social networks

Ginestra Bianconi (Northeastern Univ. – Boston, US)

Main reference K. Zhao, M. Karsai and G. Bianconi, Entropy of dynamical social networks, PlosOne (in press)

Human dynamical social networks encode information and are highly adaptive.

To characterize the information encoded in the fast dynamics of social interactions, we introduce the entropy of dynamical social networks. By analysing a large dataset of phone-call interactions we show evidence that the dynamical social network has an entropy that depends on the time of the day in a typical week-day. Moreover we show evidence for adaptability of human j social behavior showing data on duration of phone-call interactions that significantly deviates from the statistics of duration of face-to-face interactions.

This adaptability of behavior corresponds to a different information content of the dynamics of social human interactions. We quantify this information by the use of the entropy of dynamical networks on realistic models of social interactions.

3.2 Bistability Through Triad Closure

Des Higham (The University of Strathclyde – Glasgow, GB)

Joint work of Higham, Des; Grindrod, Peter; Parsons, Mark;
Main reference P. Grindrof P, D.J. Higham, M.C. Parsons, "Bistable evolving networks," University of Strathclyde, Maths and Stats Report #6, 2011.
URL http://www.mathstat.strath.ac.uk/downloads/publications/6techbis-11t.pdf

I will describe a new dynamic network model that incorporates the concept of triad closure. This feature, where new friendships are more likely to emerge between those who have more current friends in common, has been observed empirically in the social sciences. Using a discrete time Markov chain setting, mean field analysis reveals that two different types of long term behavior are possible—the system may evolve either into a sparse or a highly triangulated regime, depending on the initial data and the microscale details. Computer simulations confirm this bistability phenomenon.

3.3 The network position of MRSA risk wards in a hospital system

Petter Holme (University of Umeå, SE)

We study a dataset of patient flow in a regional Swedish hospital system. We can follow how about 300,000 patients move between 8507 wards over about ten years. The dataset also reveals when a patient test positive with Methycillin Resistant Staphylococcus aureus (MRSA).

To simplify the complex flow of patients, we represent it as a network of wards where two wards are connected if they a patient moves from one ward to the other without visiting a third ward in between. From this network we characterize the typical network position of wards with a high prevalence of MRSA, and the how the patient's location in the network changes upon testing positive with MRSA. We note that the many types of network centrality measures are positively, albeit weakly, correlated with the average prevalence of MRSA at a ward. On average, wards with medium values of these centrality measures have the highest average prevalence. We see a weak effect of the hospital system's response to the patient testing positive—after testing positive the patient move to wards with lower degree (number of links to other wards) and longer average durations of stay. One can use ward networks to aid the discovery of potential hot spots for MRSA epidemics. However, this cannot be done very effectively because even though the hospital system is administratively organized in a hierarchical way, the patient flow is too random to make the network structure a strong predictor of MRSA epidemics.

3.4 Suppression and explosive phase transitions

Byungnam Kahng (Seoul National University, KR)

When dynamics of networks proceeds under the suppression of formation of a given target pattern, the pattern can be formed eventually at a certain transition point. This formation can occur drastically, and exhibits a discontinuous (explosive) phase transition. I show some examples of such explosive transitions in percolation transition, data packet transition and epidemic spreading models.

References

3.5 Components in time-varying graphs

Vincenzo Nicosia (Università di Catania, IT)

Joint work of Nicosia, Vincenzo; Tang, John; Musolesi, Mirco; Russo, Giovanni; Mascolo, Cecilia; Latora, Vito
Main reference V. Nicosia, J. Tang, M. Musolesi, G. Russo, C. Mascolo, V.Latora, "Components in time-varying graphs," under review on Physical Review E, arXiv:1106.2134v2 [physics.soc-ph]
URL http://arxiv.org/abs/1106.2134v2

Real complex systems are inherently time-varying. Thanks to new communication systems and novel technologies, it is today possible to produce and analyze social and biological networks with detailed information on the time of occurrence and duration of each link.

However, standard graph metrics introduced so far in complex network theory are mainly suited for static graphs, i.e., graphs in which the links do not change over time, or graphs built from time-varying systems by aggregating all the links as if they were concurrent in time. In this paper, we extend the notion of connectedness, and the definitions of node and graph components, to the case of *time-varying graphs*, which are represented as time-ordered sequences of graphs defined over a fixed set of nodes. We show that the problem of finding strongly connected components in a time-varying graph can be mapped into the problem of discovering the maximal-cliques in an opportunely constructed static graph, which we name the *affine graph*. It is therefore a NP-complete problem. As a practical example, we

have performed a temporal component analysis of time-varying graphs constructed from three data sets of human interactions. The results show that taking time into account in the definition of graph components allows to capture important features of real systems. In particular, we observe a large variability in the size of node temporal in- and out-components. This is due to intrinsic fluctuations in the activity patterns of individuals, which cannot be detected by static graph analysis.

3.6 A tale of many cities: universal patterns in human urban mobility

Anastasios Noulas (University of Cambridge, GB)

Main reference A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, C. Mascolo, "A tale of many cities: universal patterns in human urban mobility," arXiv:1108.5355v4 [physics.soc-ph]
URL http://arxiv.org/abs/1108.5355v4

The advent of geographic online social networks such as Foursquare, where users voluntarily signal their current location, opens the door to powerful studies on human movement. In particular the fine granularity of the location data, with GPS accuracy down to 10 meters, and the worldwide scale of Foursquare adoption are unprecedented. In this paper we study urban mobility patterns of people in several metropolitan cities around the globe by analyzing a large set of Foursquare users. Surprisingly, while there are variations in human movement in different cities, our analysis shows that those are predominantly due to different distributions of places across different urban environments.

Moreover, a universal law for human mobility is identified, which isolates as a key component the rank-distance, factoring in the number of places between origin and destination, rather than pure physical distance, as considered in some previous works. Building on our findings, we also show how a rank-based movement model accurately captures real human movements in different cities.

Our results shed new light on the driving factors of urban human mobility, with potential applications for urban planning, location-based advertisement and even social studies.

3.7 Personality and Language in Social Media

Daniele Quercia (University of Cambridge, GB)

In Twitter, we tested whether users can be reduced to look-alike nodes (as most of the spreading models would assume) or, instead, whether they show individual differences that impact their popularity and influence. Again, one aspect that may differentiate users is their character and personality. For 335 users, we gather personality data, analyze it, and find that both popular users and influentials are extroverts and emotionally stable (low in the trait of Neuroticism). Also, since it has been shown that personality is linked to the use of language (which is unobtrusively observable in tweets), we carry out a study of tweets and show that popular and influential users linguistically structure their tweets in specific ways.

3.8 Geography of Personality

Jason Rentfrow (University of Cambridge, GB)

Everybody knows that New Yorkers are outspoken, neurotic, and always in a hurry, and that Texans are slow-talking, friendly, and proud members of the National Rifle Association. Obviously such characterizations are nothing more than stereotypes, but they raise the question of whether there are psychological differences across the US. Recent research suggests there are statewide differences in personality and that those differences are linked to a host of important social indicators. However, that work was based on one sample so the robustness of the findings is unclear. Using data from over 1.5 million respondents from five independent samples, I examined the reliability and validity of state-level personality. Analyses of the convergent validity of the state-level personality scores revealed a considerable level of convergence for each of the Big Five personality domains, with Conscientiousness displaying the least and Openness displaying the most convergence across samples. Consistent patterns of relationships across samples were observed between the state-level personality domains and conceptually relevant social indicators. For example, state-level Agreeableness was negatively related to rates of violent crime and positively related to community involvement; state-level Neuroticism was negatively related to psychological well-being and positively related to rates of cancer and mental illness; and state-level Openness was negatively related to votes cast for conservative politicians and positively related to markers of cultural diversity. Overall, these findings indicate that state-level personality is a robust and stable construct.

3.9 Distance Matters: Socio-spatial Properties of Online Social Networks

Salvatore Scellato (University of Cambridge, GB)

Joint work of Scellato, Salvatore; Noulas, Anastasios; Lambiotte, Renaud; Mascolo, Cecilia;
Main reference S. Scellato, A. Noulas, R. Lambiotte, C. Mascolo, "Socio-spatial Properties of Online Location-based Social Networks," in Proc. of Fifth Int'l AAAI Conference on Weblogs and Social Media (ICWSM'11), 2011.
URL http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2751

The spatial structure of large-scale online social networks has been largely unaccessible due to the lack of available and accurate data about people's location. However, with the recent surging popularity of location-based social services, data about the geographic position of users have been available for the first time, together with information on their online social connections.

In this work we present a comprehensive study of the spatial properties of the social networks arising among users of three main popular online location-based services. We observe robust universal features across them: while all networks exhibit about 40% of links below 100 km, we further discover strong heterogeneity across users, with different characteristic spatial lengths of interaction across both their social ties and social triads. Our results constitute the first large-scale study to unravel the socio-spatial properties of online location-based social networks.

3.10 Understanding mobility in a social petri dish

Roberta Sinatra (Università di Catania, IT)

Joint work of Szell, Michael; Sinatra, Roberta; Petri, Giovanni; Thurner, Stefan; Latora, Vito

Despite the recent availability of large data sets on human movements, a full understanding of the rules governing motion within social systems is still missing, due to incomplete information on the socio-economic factors and to often limited spatio-temporal resolutions. Here we study an entire society of individuals, the players of an online-game, with complete information on their movements in a network-shaped universe and on their social and economic interactions. Such a "socio-economic laboratory" allows to unveil the intricate interplay of spatial constraints, social and economic factors, and patterns of mobility. We find that the motion of individuals is not only constrained by physical distances, but also strongly shaped by the presence of socio-economic areas. These regions can be recovered perfectly by community detection methods solely based on the measured human dynamics. Moreover, we uncover that long-term memory in the time-order of visited locations is the essential ingredient for modeling the trajectories.

3.11 Temporal metrics and applications in real networks

John Tang (University of Cambridge, GB)

Joint work of Tang, John; Mascolo, Cecilia; Musolesi, Mirco; Latora, Vito; Nicosia, Vincenzo; Scellato, Salvatore

The study of important nodes in social and technological networks is an important research question, however the current state-of-the-art is based on static or aggregated model of the network topology. We argue that dynamically evolving network topologies are inherent in many systems, including real online social and technological networks: fortunately the nature of these systems is such that they allow the gathering of large quantities of fine-grained temporal data on interactions amongst the network members.

In this talk we shall present a temporal graph model and reformalise the concepts of shortest paths taking into account time information such as duration, frequency and time order. From this we propose novel temporal centrality metrics, namely closeness and betweenness, which take into account such dynamic interactions over time. These metrics can be applied to a large variety of dynamic networks, including mobile networks, online social networks, and in general, for the study of human interactions. In particular, using a real corporate email and human contact dataset we evaluate the important individuals selected by means of static and temporal analysis taking two perspectives: firstly, from a semantic level, we investigate their corporate role in the organisation; and secondly, from a dynamic process point of view, we devise two short-range mobile malware containments schemes based on blocking (using betweenness) and opportunistic patch spreading (using closeness). We find that temporal analysis provides a better understanding of dynamic processes and a more accurate identification of important nodes compared to traditional static methods.

4 Working Groups

During the seminar the participants split in three working groups, centered around three broad discussion themes. In the following paragraphs there is a summary of the discussion held by each working group.

- **Access to data** There is an important need to acquire large datasets depicting social interactions between individuals. However, such information is almost always controlled by large online companies or mobile phone operators: thus, it is often hard for academic researchers to get access to this kind of data, hampering research efforts. An important point is that anonymization techniques for such datasets are not mature enough to ensure no sensitive personal data can be extracted from dataset: thus, companies are not willing to share information. A potential solution might be to find new ways of collaborating with companies, so that company data can be used by academic researchers without publicly sharing any user information.
- **New models** Even though the temporal evolution and the dynamics of social networks have been largely studied, a new aspect of user behavior is increasingly becoming connected to social interactions: mobility over geographic space. Hence, new models are needed to merge together information about how users interact, when they do so and where they go. While social network structure and user mobility have been separately investigated, with important outcomes, new efforts must be spent to define new models which include both aspects.
- **Practical applications** The practical outcomes of research efforts can be roughly divided in two broad classes: models of user behavior and design of new systems and applications. Companies are more suitable to pursue the latter class, as they benefit from larger amount of resources and data than academic entities. However, academic research has an invaluable advantage in understanding and modeling the universal factors behind user behavior, providing insights that can fuel further research, both in academic and company settings. A mutual and lively exchange of ideas and results between the two fields is hence mandatory.

5 Panel Discussion

The seminar hosted a panel discussion, which centered around four main themes.

- **Theoretical challenges** The analytical challenges arising from recent results regard how to deal with time and space. In particular, the effect of space on user behavior must be better understood, exploiting the huge literature of geography studies. At the same time, dealing with how these systems change over time requires extending models and techniques to include time scales. In general, these two aspects must be intimately combined to provide new spatio-temporal models which also include social behavior.
- **Application challenges** The growing popularity of online services requires users to maintain several versions of their identity, depending on the services they are using. This generates the problem of managing the connection between the real, offline identity and the multiple online identities, protecting privacy and personal information at the same time. In particular, as personal data become scattered across different companies, new mechanisms to manage trust and reputation must be designed, tested and implemented.

- **Future directions** Even though it might be difficult to have an important impact as an academic researcher, since large-scale companies have more resources to solve real-world problems, there is still a lot of value in academic research output. In particular, academic efforts should try to focus on tasks that are usually outside the scope of company laboratories, such as theoretical models and interdisciplinary studies.
- **A psychology perspective** Even though a vast number of works and studies consider social networks as large-scale systems where the nature of individual nodes can be neglected, all nodes are not equal, since they represent human entities. Thus, there are two important consequences: the introduction of psychological and sociological theories can further help research efforts and, on the other hand, large-scale data analysis can help to study psychological problems. At the end, an important result could be to connect user behavioral patterns with personality aspects.

Participants

- Ginestra Bianconi
Northeastern Univ. – Boston, US
- Vincent D. Blondel
UC Louvain-la-Neuve, BE
- Sonja Buchegger
KTH Stockholm, SE
- Krishna P. Gummadi
MPI für Software Systeme – Saarbrücken, DE
- Des Higham
The University of Strathclyde – Glasgow, GB
- Petter Holme
University of Umeå, SE
- Byungnam Kahng
Seoul National University, KR
- Renaud Lambiotte
Imperial College London, GB
- Vito Latora
Università di Catania, IT
- Cecilia Mascolo
University of Cambridge, GB
- Alan Mislove
Northeastern Univ. – Boston, US
- Mirco Musolesi
University of Birmingham, GB
- Vincenzo Nicosia
Università di Catania, IT
- Anastasios Noulas
University of Cambridge, GB
- Eamonn O'Neill
University of Bath, GB
- Alessandro Panconesi
University of Rome "La Sapienza", IT
- Daniele Quercia
University of Cambridge, GB
- Jason Rentfrow
University of Cambridge, GB
- Stefan Saroiu
Microsoft Res. – Redmond, US
- Salvatore Scellato
University of Cambridge, GB
- Roberta Sinatra
Università di Catania, IT
- Michael Sirivianos
Telefonica Res. – Barcelona, ES
- Thorsten Strufe
TU Darmstadt, DE
- John Tang
University of Cambridge, GB
- Walter Willlinger
AT&T Res. – Florham Park, US

Report from Dagstuhl Seminar 11461

Coding Theory

Edited by

Joachim Rosenthal[1], M. Amin Shokrollahi[2], and Judy Walker[3]

1 **Universität Zürich, CH,** `rosenthal@math.uzh.ch`
2 **EPFL – Lausanne, CH,** `amin.shokrollahi@epfl.ch`
3 **University of Nebraska – Lincoln, US,** `jwalker7@math.unl.edu`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 11461 "Coding Theory". A (channel) code is typically a set of vectors of the same length n over a finite alphabet Σ. By choosing a fixed codebook, binary strings of appropriate length are injectively mapped into the elements of the code. These elements are then transmitted over a communications channel which induces errors on the codeword. Depending on how well the original code is designed, and which algorithms are used, the result of this transmission and attempts to recover the original vector after transmission can be anywhere between disastrous to excellent. Coding theory is all about the design of excellent codes as a function of the communications channel, and the design of efficient algorithms for choosing the codebook vectors, and more importantly, for recovering the original vector after transmission. As such, successful design of codes requires knowledge and tools in a number of areas such as combinatorics, algorithms design, probability theory and complexity theory, to name a few. The purpose of this workshop is to bring together researchers in the field to discuss recent theoretical advances in algebraic coding, codes on graphs, and network coding, as well as new and emerging applications of coding methods to real-world problems.

Seminar 13.–18. November, 2011 – www.dagstuhl.de/11461
1998 ACM Subject Classification E.4 Coding and Information Theory, I.1.2 Algorithms.
Keywords and phrases Algebraic coding theory, complexity theory, cryptography, graph theory, graph based codes, information theory, randomized algorithms, networking, data distribution.
Digital Object Identifier 10.4230/DagRep.1.11.50
Edited in cooperation with Ghid Maatouk, Masoud Alipour

1 Executive Summary

Joachim Rosenthal
M. Amin Shokrollahi
Judy Walker

This workshop brought together 42 researchers in key areas of coding theory. The seminar had a strong emphasis on interaction and collaboration among researchers. This goal was made clear in a series of five-minute talks by all seminar participants, where they briefly described their research interests and/or described a problem that they were currently working on. The rest of the time was dedicated to more in-depth talks by a selected number of researchers, but the lecture time was kept reasonably short so that long stretches of time were available for seminar participants who wished to discuss further and collaborate in small groups.

The talks fell into several broad categories, the most prominent of which was algebraic coding theory. Algebraic coding theory primarily investigates codes obtained from algebraic constructions. Prime examples of this area of coding theory are codes from algebraic geometry and codes obtained from algebraically constructed expander graphs. This discipline is almost as old as coding theory itself, and has attracted (and continues to attract) some of the brightest minds in the field. Among the most exciting advances in this field in recent years has been the invention of list-decoding algorithms for various classes of algebraic codes; this area began with a landmark paper by Sudan that proposed an algebraic list-decoding scheme for Reed-Solomon codes. List decoding algorithms yield for a received word a short list of codewords that have at most a given distance τ to the received word. The size of the list depends on the distance τ. The methods in this field are mostly algebraic and make use of various properties of multivariate polynomials, or more generally, the properties of "well-behaved" functions in the function field of an irreducible variety. Methods from algebraic geometry are very important in this area. On the computational side the field naturally embeds in the theory of Gröbner bases. There are emerging relationships between this area and codes on graphs, the leading question being whether it is possible to match the superior performance of graph-based codes with list-decoding algorithms, or at least with algorithms that are derived from list-decoding algorithms.
Several talks covered recent advances and current research in some of the most notable questions that relate to algebraic constructions of codes, namely, list decoding; Berlekamp-Massey-like algorithms (for decoding, list-decoding, and Gröbner bases computation); elliptic curve methods and bent and hyperbent functions; rank-metric codes; bounds on codes (semidefinite programming bounds, BCH-like bounds); pivot distributions; properties of specific code constructions such as cyclic orbit codes, classes of self-dual codes, and codes obtained from generalized concatenation; and pseudocodewords, which straddle the boundary between algebraic analysis of codes and message-passing decoding of codes on graphs.

The second area that was covered in the workshop was that of codes on graphs. First proposed by Gallager in the 1950s, the subject of codes on graphs has experienced a huge revival over the past 10–15 years due to the fact that these codes have been shown to have capacity-achieving properties on some channels, and capacity-approaching properties on others. One of the most prominent examples in this area is that of the class of low density parity check (LDPC) codes, which are constructed from sparse bipartite graphs. The sparsity of the graph provides methods for construction of low complexity encoders and decoders. This area is a perfect nurturing ground for cross-fertilization of ideas between computer science, electrical engineering and mathematics. On the engineering side, the amazing simulation results have led some to declare the channel coding problem solved. The main proofs of asymptotic performance of these codes originated in the theoretical computer science community. Despite serious activities in this field many questions remain, related to the performance of graph-based codes and design of good LDPC codes. Several speakers discussed topics related to graph-based codes with applications beyond this field. The topics covered included design of sparse-graph codes for cooperation in communication networks; a new class of iterative decoders for LDPC codes; classes of tail-biting treillises; analysis of treillis pseudocodewords distributions; and bounds on the performance of quantum LDPC codes and their applications to percolations on graphs.

Network coding was another main area covered by the seminar. Network coding theory is concerned with the encoding and transmission of information where there may be many

information sources and possibly many receivers. This is a very new area of coding theory and has been around only since 1999. The topic is somewhat far removed from the channel coding problem that Shannon proposed in 1948 and that the other areas of coding theory described in this proposal address. Indeed, there is not yet an agreed-upon formulation of the network coding problem. But there is a notion of linear network coding, and Li, Yeung and Cai showed in 2003 that it is possible to achieve so-called network capacity using linear codes alone. From that basis, much more algebra has been fruitfully introduced into the area. For example, in 2003, Kötter and Médard showed that the linear network codes that achieve a set throughput on a given network are precisely described by the points in a certain algebraic variety associated to the network. The Kötter-Médard approach, of course, relies on knowledge of the network in question. In practice, the network is typically unknown and often continually changing (dynamic). In 2008, Kötter and Kschischang considered random network coding and presented a framework in which the problem of network design is separated from that of code design. The idea is to assume that the network source simultaneously injects k linearly independent vectors from some vector space W into the network. These vectors are combined in various ways and sent through the network, so that the sink receives some linear combinations of them. The mathematical object that is invariant during transmission is the subspace of W spanned by the original k vectors, and so it is natural to consider a code in this context to be a subset of the set $\mathcal{P}(W)$ of all subspaces of W. Since $\mathcal{P}(W)$ is a metric space, the questions of code construction and optimality arise. Some approaches from classical coding theory can therefore be adopted in a fairly straight-forward manner, but deep questions remain, some of which were addressed by seminar speakers. One talk addressed the design and analysis of good end-to-end error-control codes in linear network coding; another topic was the analysis of several strategies for content distribution over network coded systems.

No workshop on such a practical topic as coding theory is complete without the mention of new and emerging applications. Indeed, several of the covered topics had a decidedly practical flavor, as some leading experts in applied coding theory, with an extensive mathematical background, reported about their work and provided insights into the directions the field will be taking in the coming years. Some of the most striking applications of coding-theoretical techniques to practical problems that were discussed included explicit constructions of regenerating codes for distributed storage; using network coding techniques to increase throughput in content distribution; and the design of iterative decoders for LDPC codes with better error floors than the traditional belief propagation decoders.

2 Table of Contents

3 Overview of Talks

3.1 Coding for Relay Channels

Iryna Andriyanova (Université de Cergy-Pontoise/ENSEA/CNRS, FR)

We present two methods of designing efficient sparse-graph codes for cooperation. The first method is an extension of the design procedure for a point-to-point communication and uses good error-correcting codes (ex. LDPC) both at the source and at the relay ends. The second method takes the advantage of the multi-point transmission model and uses good error-detecting codes (ex. LDGM) along with a block-Markov encoding at the source end. We show that both of the design procedures have comparable performances asymptotically and may achieve the capacity of the relay channel.

3.2 List decoding and weighted reconstruction

Daniel Augot (École Polytechnique – Palaiseau)

In list decoding, a very important variant of list decoding of Reed-Solomon codes is the weighted polynomial reconstruction problem, as introduced by Guruswami and Sudan, which was later used for soft decoding of Reed-Solomon codes by Koetter and Vardy. In that context, a multiplicity matrix is built, which consists in interpolation constraints for finding the bivariate interpolation polynomial which is such that it admits linear factors corresponding to codewords. We investigate the simple case of a symmetric channel, where we also assume that the interpolation matrix is column symmetric. In that case, we identify the problem as the sum of simple list decoding problems. This can be described in an ideal-theoretical way, where the looked for ideal is the intersection of explicit ideals. We discuss how the linear factors of the simpler interpolation polynomials given by these ideals relate to the solution of the global interpolation problem, in an analogy with ideal-variety correspondence.

3.3 Semidefinite programming bounds for projective codes

Christine Bachoc (Université Bordeaux, FR)

Joint work of Bachoc, Christine; Passuello, Alberto; Vallentin, Frank

A projective code C is a subset of the set $P_q(n)$ of linear subspaces of $\mathbb{F}_q^n$. We consider on $P_q(n)$ the subspace distance defined by $d(U,V) = \dim(U)+\dim(V)-2\dim(U\cap V)$. Following [2], this distance is relevant in the framework of network coding. We consider in this context the classical coding theoretic problem of giving upper estimates for the maximal number $A_q(n,d)$ of elements of a projective code with given minimal distance d. If the classical bounds (anticode, singleton, Johnson, Delsarte, etc..) have been extended in an appropriate way to the case of subspaces of equal dimension, it appears that they cannot apply so easily

to general projective codes. This is essentially due to the fact that the Grassmann space $G_q(n,k)$ is homogeneous under the action of the linear group $Gl(n,q)$ while the projective space is not. For general projective codes, only T. Etzion and A. Vardy [4] have provided an interesting bound. Their bound is expressed as the optimal value of a linear program and can be seen as an analog of the sphere packing bound.

In this talk, we present a bound for $A_q(n,d)$ which is the optimal value of a semidefinite program of size polynomial in n, obtained by the symmetrization of a general program, which itself can be viewed as the Lovász theta number of a certain graph [3]. This result is essentially a direct application of the method explained in [1] in a more general context. Numerical computations for $n \leq 13$ show that this bound is slightly better than the Etzion-Vardy bound.

References

1 C. Bachoc, D. Gijswijt, A. Schrijver and F. Vallentin, *Invariant semidefinite programs*, to appear in the Handbook on Semidefinite, Conic and Polynomial Optimization (M.F. Anjos, J.B. Lasserre (ed.)).

2 R. Koetter and F. R. Kschischang, *Coding for errors and erasures in random network coding*, IEEE Transactions on Information Theory, Volume 54, Issue 8, August 2008, pp. 3579–3591.

3 L. Lovász, *On the Shannon capacity of a graph*, IEEE Trans. Inform. Theory vol. 25 (1979), pp. 1–5.

4 T. Etzion and A. Vardy, *Error-correcting codes in projective space*, IEEE Trans. Inform. Theory, vol. 57, no. 2, pp. 1165–1173, February 2011.

3.4 A Berlekamp-Massey like decoding algorithm

Martin Bossert (Universität Ulm)

An algorithm to determine the error locator polynomial for decoding Reed-Solomon codes is presented. The novel approach uses only properties of linear codes for the proof. The possible extension to the case of list decoding will be explained.

3.5 Trellises and Pseudocodewords

David Conti (University College – Dublin)

Joint work of Conti, David; Boston, Nigel

A central paradigm in modern coding theory is to represent codes by special graphs that allow for efficient iterative decoding algorithms. Amongst the most notable of such graphs are *trellises* ([7]). These are dynamical representation of codes that are important both for decoding purposes and for their rich theory which gives combinatorial insight into codes ([2, 6, 7]). Iterative decoding performance on a trellis is affected by *pseudocodewords*, in particular those with low *pseudoweight* ([4, 5]), while decoding complexity grows with trellis size ([7]). To make a trellis smaller one needs to merge vertices, but unfortunately this could

create extra cycles yielding bad pseudocodewords. Thus it is important to study trellises along with their pseudocodewords.

In this talk we will first take a stroll through trellises and their pseudocodewords by discussing concretely the basics and some key questions. We will then show how an algebraic framework can be developed to help us studying trellis pseudocodewords distributions, bringing into the picture recurrence sequences, symmetric functions, and invariant theory ([3]). This can then be applied to give a first partial answer to a conjecture on the minimal 16 state trellis for the extended binary Golay code ([1, 2]).

References

1 N. Boston, *A multivariate weight enumerator for tail-biting trellis pseudocodewords*, to appear in Proceedings of "Workshop on Algebra, Combinatorics and Dynamics" Belfast, August 2009, Journal of Mathematical Sciences, Springer

2 A. R. Calderbank, G. D. Forney, Jr., A. Vardy, *Minimal tail-biting trellises: The Golay code and more*, IEEE Trans. Inform. Theory 45 (5) (1999) 1435–1455

3 D. Conti, N. Boston, *Matrix representations of trellises and enumerating trellis pseudocodewords*, Accepted to appear in Allerton 2011 proceedings

4 G. D. Forney Jr., R. Koetter, F. R. Kschischang, A. Reznik, *On the effective weights of pseudocodewords for codes defined on graphs with cycles*, in: Codes, Systems and Graphical Models, Springer, New York, 2001, 101–112

5 G. B. Horn, *Iterative decoding and pseudocodewords*, Ph.D. Thesis, California Institute of Technology, Pasadena, California, 1999

6 R. Koetter, A. Vardy, *The structure of tail-biting trellises: minimality and basic principles*, IEEE Trans. Inform. Theory 49 (9) (2003) 2081–2105

7 A. Vardy, *Trellis structure of codes*, in: Handbook of Coding Theory, Volume 2, Elsevier Science, 1998

3.6 Pivot distributions and Weierstrass nongaps

Iwan M. Duursma (University of Illinois – Urbana-Champaign)

We show how various properties of a linear code are captured by the collection of its dimension-length profiles. The dimension-length profile of a matrix describes the column positions of the pivots after the matrix is brought into row echelon form. The collection of dimension-length profiles describes the distribution of the pivots considered over all possible column permutations of the matrix. A code is MDS if and only if for every permutation of the columns in the generator matrix the pivot columns are the leading columns. The zeta function of a code that we introduced in previous work will be used to describe the deviation from this distribution when the code is not MDS. We will point out how the zeta function is the exact analogue of the zeta function for curves over a finite field and how for curves it describes the distribution of Weierstrass nongaps.

3.7 Classes of Tail-Biting Trellises

Heide Gluesing-Luerssen (University of Kentucky, US)

Joint work of Gluesing-Luerssen, Heide; Forney, G. David Jr.

Tail-biting trellises are – just like the ubiquitous Tanner graphs – instances of normal graphs for codes. More specifically, both form graphs with cycles. While the theory of cycle-free graphs is by now well understood with respect to minimal realization as well as decoding, many questions are still open for graphs with cycles. Yet, these graphs form an important class in coding theory and practice since it has long been known that realizations on graphs with cycles may allow for much more powerful iterative decoding algorithms than cycle-free graphs.

In this talk we will focus on tail-biting trellises: these are graphs in which the constraint codes form a single cycle. We will aim at identifying particularly useful tail-biting trellises. It is well-known that the complexity of decoding algorithms such as the sum-product algorithm depends on the size of the state spaces as well as constraint codes, and thus – not surprisingly – the construction and understanding of minimal trellises is paramount to the theory of tail-biting trellises. However, it turns out that all meaningful concepts of complexity measures for tail-biting trellises lead to different orderings, all of which are only partial, and thus to distinct notions of minimality. In particular, none of these notions leads to a unique minimal trellis realization (in the specified sense) for a given code.

As a consequence, the question of classifying good trellises becomes a crucial yet non-trivial step in the theory of tail-biting trellises. A major tool for identifying structural properties is the normal graph dualization introduced by Forney (2001). Indeed, it turns out that the dual graph may reveal trellis properties that are not immediately identifiable from the primal trellis. In this talk, we will show several such properties and relate them to more obvious properties of the dual trellis. Among other things, we will present system-theoretically meaningful notions of controllability and observability. As to be expected they are mutually dual. The process of investigating a trellis and its dual in parallel will also lead to a constructive reduction procedure for certain trellises.

3.8 A BCH-Bound for Ring-Linear Coding

Marcus Greferath (University College – Dublin, IE)

Joint work of Greferath, Marcus; Conti, David; Walker, Judy; Zumbrägel, Jens

Ring-Linear Coding started enjoying increased attention at the beginning of the last decade of the previous century, when it was discovered that certain non-linear binary block codes of high quality could be linearized by a change of alphabet. These new code versions were linear over a ring, rather than a field, and also the metric on this ring had to be different from the Hamming metric to fully reflect their error correction properties.

Since those days a lot of research has been devoted to ring-linear coding. On the structural side there was a great success settling the question which class of rings (or modules) is the most appropriate for ring-linear coding, and allows for possibly most foundational results of the traditional finite-field based theory.

On the constructional side we observe less success in providing families or sporadic examples of codes which outperform their finite-field siblings. This is in part due to the fact that up to now there have not been proved any assertive bounds like the BCH bound that allows to construct codes of a desired minimum distance.

The present talk's goal is to tackle this problem. We will show that a BCH bound can be shown for *any* finite ring R as long as we observe the expected restriction that the length n of the code in question is an invertible element in R. As a by-product we obtain a very nice new characterization of the class of all Frobenius rings. This is given as follows:

Theorem: For a finite ring R the following are equivalent.

(a) R is a Frobenius ring.
(b) There is a homomorphism $\chi : (R, +) \longrightarrow (\mathbb{C}^\times, \cdot)$, such that

$$\sum_{y \in I} \chi(y) \;=\; \begin{cases} 1 & : \quad I = \{0\}, \\ 0 & : \quad \text{otherwise.} \end{cases}$$

(c) For every ring S there exists an extension T and a homomorphism $\chi : (R, +) \longrightarrow (T^\times \cap Z(T), \cdot)$ that satisfies the condition under **(b)**.

3.9 Linear-algebraic list decoding and subspace-evasive sets

Venkatesan Guruswami (Carnegie Mellon University – Pittsburgh)

Joint work of Guruswami, Venkatesan; Wang, Carol; Xing, Chaoping

This talk described a simple linear-algebraic approach that is surprisingly effective in list decoding variants of Reed-Solomon codes from an error fraction approaching the information-theoretically maximum limit of 1-R where R is the rate of the code.

The algorithm can be thought as a higher-dimensional version of the Welch-Berlekamp decoder. Similar to earlier algebraic list-decoders, the algorithm consists of two steps: multivariate polynomial interpolation (albeit with no need for "multiplicities") followed by a "root-finding" step to determine candidate close-by codewords. Both steps amount to solving (certain structured) linear systems and can be performed efficiently. Further, the candidate solutions are confined to a low-dimensional subspace. Pruning this subspace to find close-by codewords, though polynomial time, is the theoretically most expensive step of the algorithm. However, in practice one expects that the subspace will consist of a unique (or very few) candidate messages.

By pre-coding the messages to belong to a large "subspace-evasive set", i.e., a subset that has small intersection with every low-dimensional subspace, we can reduce the proven bound on list size (after pruning the candidate subspace) to a constant. Such subspace-evasive sets are interesting pseudorandom objects in their own right. I also stated state the outcome of recent generalizations of this approach to folded algebraic-geometric codes, leading to efficiently list-decodable codes with simultaneously near-optimal rate, list size, and alphabet size.

3.10 Bent functions and related topics

Tor Helleseth (University of Bergen, NO)

Joint work of Helleseth, Tor; Carlet, Claude; Kholosha, Alexander; Mesnager, Sihem

Bent functions are Boolean functions that have maximal distance to all affine functions. In the first part of the talk we gave an overview of classical constructions of bent functions as well as some recent constructions that relates the known families of Niho bent functions to o-polynomials and hyperovals in finite geometry. The corresponding o-polynomials of some of the Niho bent functions were found and the duals of some of these bent functions were given along with some open problems.

The last part of the talk presented an overview of recent results on generalized (nonbinary) bent functions.

3.11 Network Error-Control Coding

Frank R. Kschischang (University of Toronto)

Joint work of Kschischang, Frank R.; Silva, Danilo; Koetter, Ralf

The problem of error-control coding in linear network coding was reviewed. In the case of *coherent* network coding, in which the source-destination transfer matrix is known in advance, it was shown that optimal end-to-end error-control against adversarial errors can be achieved using rank-metric codes. In the case of *noncoherent* network coding, in which the source-destination transfer matrix is random and unknown to transmitter and receiver, it was shown that subspace codes, consisting of codebooks of vector-spaces well-separated according to an appropriate metric, are a good choice for end-to-end error-control against adversarial errors. Simple bounds on such codes were reviewed, and a *lifting* construction described, allowing for the construction of good subspace codes from rank-metric codes. Finally, a random (non-adversarial) noncoherent matrix channel was described. Capacity bounds were given, and a capacity-approaching transmission scheme involving *channel sounding* and *error-trapping* was described, for which the corresponding decoder is simply Gauss-Jordan elimination.

3.12 Interpolation, parametrization and iterative Grobner methods

Margreta Kuijper (The University of Melbourne, AU)

In this talk we consider minimal interpolation and advocate iterative algorithms that construct a minimal Grobner basis at each step. Already in 1995 Fitzpatrick introduced such an algorithm, resembling the Berlekamp-Massey algorithm. In the talk we explicitly use the "predictable leading monomial" property and consider parametric list decoding of Reed-Solomon codes as well as iterative shortest LFSR construction over finite rings.

3.13 Explicit Constructions of Regenerating Codes for Distributed Storage

P. Vijay Kumar (Indian Inst. of Science – Bangalore, IN)

Joint work of Kumar, P. Vijay; Shah, Nihar; Rashmi, K.V.; Ramachandran, K.

Regenerating codes are a class of distributed storage codes that optimally trade the bandwidth needed for repair of a failed node with the amount of data stored per node of the network. An [n; k; d] regenerating code permits the data to be recovered by connecting to any k of the n nodes in the network, while requiring that repair of a failed node be made possible by connecting to any d nodes. The amount of data downloaded for repair is typically much smaller as compared to the size of the source data. In this talk, we present explicit and optimal constructions of regenerating codes.

3.14 Polar codes as generalized concatenation

Simon Litsyn (Tel Aviv University, IL)

Polar construction of codes has attracted a plenty of attention recently since it provides a class of codes provably achieving capacity of symmetric memory-less channels and having low complexity of implementation.

In the talk we show that the original approach due to Arikan can be seen as recursively applied generalized concatenated construction with the inner codes having length 2. This suggests several natural extensions, e.g. using longer inner nested codes and mixed alphabet kernels. These modifications yield a significant improvement in performance of the considered codes.

3.15 Decoding cyclic orbit codes and the discrete logarithm problem

Felice Manganiello (University of Toronto, CA)

Cyclic orbit codes are a subfamily of orbit codes for linear random network codes. Finding a decoding algorithm for these family of codes is still an open problem. In this seminar we show the connection between decoding cyclic orbit codes and a problem called rank discrete logarithm problem. We also present a membership criterion for some particular cyclic orbit codes which is based on solving the rank discrete logarithm problem.

3.16 An efficient characterization of a family of hyperbent functions with multiple trace terms

Sihem Mesnager (Université de Paris VIII, FR)

Joint work of Mesnager, Sihem; Flori, Jean-Pierre

Lisonek recently reformulated the characterization of Charpin and Gong of a large class of hyperbent functions in terms of cardinalities of hyperelliptic curves.

In this paper, we show that such a reformulation can be naturally extended to a distinct family of functions proposed by Mesnager. Doing so, a polynomial time and space test is obtained to test the hyperbentness of functions in this family. Finally we show how this reformulation can be transformed to obtain a more efficient test.

3.17 Equivalence and Duality for Rank-Metric Codes

Katherine Morrison (University of Nebraska – Lincoln, US)

Rank-metric codes have garnered significant attention due to their applications in network coding, public-key cryptography, and data storage among other areas.

We focus on characterizing rank-metric codes that are both efficient, i.e. have high dimension, and effective at error correction, i.e. have high minimum distance. With an aim towards finding codes with a perfectly balanced trade-off, we study self-dual rank-metric codes. In particular, we seek to enumerate the equivalence classes of self-dual matrix codes of short lengths over small finite fields. Towards this end, we examine the notion of equivalence for rank-metric codes and use this to characterize the automorphism groups of a particular class of these codes known as Gabidulin codes.

3.18 Pseudocodewords from permanents

Roxana Smarandache (San Diego State University, US)

In this talk, we present some of the directions of research we were led to recently: computing or simplifying the computation of the permanent of 0-1 block matrices, studying if properties that are true for determinants hold also for permanents, defining the notion of quasipermanents for general square matrices over noncommutative rings, similar to the notion of quasideterminants, based on the reduction formulas obtained for the permanent of block matrices, and studying the properties of the "block-cofactor expansion inequality". These questions rise all from the need of information on the class of pseudocodewords with components defined by permanents.

3.19 Double Dixie Cup Unicast

Emina Soljanin (Bell Labs – Murray Hill, US)

Joint work of Soljanin, Emina; Li, Yao; Spasojevic, Predrag

In chunk based content distribution, files are fragmented at the source and the fragments (chunks) are distributed individually throughout the network. In network coded systems, instead of distributing the original file chunks, nodes send out linear combinations of the chunks they hold. This strategy brings about an overall throughput increase. For a number of practical concerns (e.g, computational complexity and synchronization), chunks are grouped into (possibly overlapping) blocks known as generations. Only chunks within the same generation are allowed to be linearly combined. A price to pay for coding only within generations is throughput reduction.

For the purpose of gaining insight into differences among these three distribution strategies, we model them as random allocation protocols, such as the classical coupon collecting and the less known double Dixie cup collecting. Since chunks represent coded information rather than indivisible coupons, we modify these allocation models accordingly. We characterize the throughput behavior in terms of the generation size, and develop strategies to improve the throughput while maintaining the benefits of coding within generations.

3.20 Finite alphabet iterative decoders

Bane Vasic (University of Arizona – Tucson, US)

We introduce a new class of finite alphabet iterative decoders for LDPC codes, and show how to choose nonlinear message update maps for column-weight-three codes so that the decoders with only seven message levels surpass belief propagation (BP) in the error floor region.

3.21 Singly-even self-dual codes with minimal shadow

Wolfgang Willems (Universität Magdeburg, DE)

Joint work of Willems, Wolfgang; Bouyuklieva, S.

There are several reasons why extremal singly-even self-dual codes are of interest. For instance, they provide connections to designs, are proposed for use in secret sharing schemes, and for length 24m+8 they perform better than doubly-even self-dual codes of the same length if the shadow is minimal. In the talk we investigate such codes with minimal shadow. For particular parameters we prove non-existence theorems and state explicit bounds whose existence was proved by Rains.

3.22 Percolation and Quantum Erasure Correction

Gilles Zémor (Université Bordeaux, FR)

Joint work of Zémor, Gilles; Delfosse, Nicolas

We show that percolation on graphs, in particular on trees, is closely related to the probability of successful decoding on the erasure channel for the class of cycle codes on graphs. We investigate what information-theoretic bounds on LDPC codes contribute to the determination of critical probabilities for percolation. Transposed to the quantum setting, we find that the same reasoning applies when cycle codes of graphs are replaced by surface codes, quantum codes built on 2-dimensional complexes. We derive new information-theoretic upper bounds on the probability of successful decoding for quantum LDPC (sparse) codes on the quantum erasure channel, and apply these bounds to obtain upper bounds on critical probabilities for percolation on tilings of the hyperbolic plane.

3.23 On the Pseudocodeword Redundancy of Binary Linear Codes

Jens Zumbrägel (University College – Dublin, IE)

Joint work of Zumbrägel, Jens; Boston, Nigel; Flanagan, Mark F.; Skachek, Vitaly

The AWGNC, BSC, and max-fractional pseudocodeword redundancies of a binary linear code are defined to be the smallest number of rows in a parity-check matrix such that the corresponding minimum pseudoweight is equal to the minimum Hamming distance of the code. We show that most codes do not have a finite pseudocodeword redundancy. Also, we provide upper bounds on the pseudocodeword redundancy for some families of codes with t-transitive automorphism group, and we report on our progress of proving bounds for the extended Golay code.

Participants

- Masoud Alipour
EPFL – Lausanne, CH
- Iryna Andriyanova
Univ. de Cergy-Pontoise/ ENSEA/CNRS, FR
- Daniel Augot
Ecole Polytechnique – Palaiseau, FR
- Christine Bachoc
Université Bordeaux, FR
- Martin Bossert
Universität Ulm, DE
- Joan Josep Climent
Universitat d'Alacant, ES
- Gerard Cohen
ENST – Paris, FR
- David Conti
University College – Dublin, IE
- Iwan M. Duursma
Univ. of Illinois – Urbana, US
- Michele Elia
Politecnico di Torino, IT
- Heide Gluesing-Luerssen
University of Kentucky, US
- Elisa Gorla
Universität Basel, CH
- Marcus Greferath
University College – Dublin, IE
- Venkatesan Guruswami
Carnegie Mellon University – Pittsburgh, US
- Kathryn Haymaker
Univ. of Nebraska – Lincoln, US
- Tor Helleseth
University of Bergen, NO
- Tom Hoholdt
Technical University of Denmark – Lyngby, DK
- Christine A. Kelley
Univ. of Nebraska – Lincoln, US
- Axel Kohnert
Universität Bayreuth, DE
- Frank R. Kschischang
University of Toronto, CA
- Margreta Kuijper
The University of Melbourne, AU
- P. Vijay Kumar
Indian Inst. of Science – Bangalore, IN
- Francoise Levy-dit-Vehel
ENSTA – Paris, FR
- Yao Li
Rutgers University – New Brunswick, US
- Simon Litsyn
Tel Aviv University, IL
- Hans-Andrea Loeliger
ETH Zentrum – Zürich, CH
- Ghid Maatouk
EPFL – Lausanne, CH
- Felice Manganiello
University of Toronto, CA
- Sihem Mesnager
Université de Paris VIII, FR
- Katherine Morrison
Univ. of Nebraska – Lincoln, US
- Joachim Rosenthal
Universität Zürich, CH
- Davide Schipani
Universität Zürich, CH
- M. Amin Shokrollahi
EPFL – Lausanne, CH
- Roxana Smarandache
San Diego State University, US
- Emina Soljanin
Bell Labs – Murray Hill, US
- Anna-Lena Trautmann
Universität Zürich, CH
- Bane Vasic
Univ. of Arizona – Tucson, US
- Judy L. Walker
Univ. of Nebraska – Lincoln, US
- Alfred Wassermann
Universität Bayreuth, DE
- Wolfgang Willems
Universität Magdeburg, DE
- Gilles Zémor
Université Bordeaux, FR
- Jens Zumbrägel
University College – Dublin, IE

Report from Dagstuhl Seminar 11471

Efficient Algorithms for Global Optimisation Methods in Computer Vision

Edited by

Andrés Bruhn[1], Thomas Pock[2], and Xue-Cheng Tai[3]

1 **Universität des Saarlandes, DE,** bruhn@mmci.uni-saarland.de
2 **TU Graz, AT,** pock@icg.tugraz.at
3 **University of Bergen, NO and Nanyang Technology University, SG,** tai@mi.uib.no

Abstract

This report documents the program and the results of Dagstuhl Seminar 11471 *Efficient Algorithms for Global Optimisation Methods in Computer Vision*, taking place November 20–25 in 2011. The focus of the seminar was to discuss the design of efficient computer vision algorithms based on global optimisation methods in the context of the *entire* design pipeline. Since there is no such conference that deals with all aspects of the design process – *modelling*, *mathematical analysis*, *numerical solvers*, and *parallelisation* – the seminar aimed at bringing together researchers from computer science and mathematics covering all four fields.

Seminar 20.– 25. November, 2011 – www.dagstuhl.de/11471
1998 ACM Subject Classification I.4 Image Processing and Computer Vision, G.1.6 Global Optimization, G.2.1 Combinatorial Algorithms, F.2.1 Analysis of Algorithms and Problem Complexity, G.1.0 Parallel Algorithms.
Keywords and phrases Computer Vision, Modelling, Mathematical Foundations, Data Structures, Efficient Algorithms, Parallel Computing
Digital Object Identifier 10.4230/DagRep.1.11.66

1 Executive Summary

Andrés Bruhn
Thomas Pock
Xue-Cheng Tai

Most of the leading algorithms in computer vision are based on global optimisation methods. Such methods compute the solution of a given problem as minimiser of suitable cost functional that penalises deviations from previously made assumptions and integrates them over the entire image domain. While their transparent modelling allows for excellent results in terms of quality for many fundamental computer vision tasks such as *motion estimation*, *stereo reconstruction*, *image restoration*, *shape matching*, and *object segmentation*, the minimisation of these cost functional often leads to optimisation problems that are mathematically challenging and computationally expensive.

In the last decade, this fact has triggered a variety of different research directions that try to satisfy the needs of an ever increasing resolution in image and video data as well as the strong real-time demands of industrial applications. These research directions can be

roughly divided into four fields that correspond to the different stages of the algorithmic design pipeline:

- *Modelling* (suitable priors, continuous vs. discrete, hybrid approaches)
- *Mathematical Analysis* (convex vs. non-convex, error bounds, well-posedness)
- *Numerical Solvers* (recent techniques, trends in numerical/combinatorial optimisation)
- *Parallelisation* (GPUs, mutli-core, cluster systems, FPGAs)

Since there are no conferences that address all four fields, the goal of this seminar was to identify and address open questions in these four fields in the context of the *entire* design pipeline. To this end, it brought together computer scientists and mathematicians from all stages. Apart from stimulating interdisciplinary research and establishing close collaborations between the different fields by scheduling plenty of time for discussions, the ultimate goal of the seminar was to develop more precise and more efficient algorithms that are conceptually well designed, mathematically well understood and from which all parts are chosen carefully such that they harmonise with each other.

Further aims of the seminar were to establish suitable benchmarks for measuring the performance of each of the stages (model precision, optimisation accuracy, numerical efficiency, parallelisability) and to derive general conceptual guidelines for the design of efficient algorithms that are applicable to a broad class of key problems in computer vision.

The seminar was conducted in a conference style, where every participant gave a talk of about 20 to 25 minutes. There was much time for extensive discussions – directly after the talks, in dedicated working groups and in the evenings. As documented by the very positive evaluation and the detailed summaries of the working groups, there was a very open and constructive atmosphere. In particular the people appreciated the integration of young researchers in the seminar schedule and the gain of new cross-disciplinary insights from talks and discussions of other participants. While the first observation reflects the fact that the field of efficient algorithms in computer vision is relatively new, the second aspect demonstrates the success of the seminar to bring together people from different communities and establish new ties between the fields. At the moment that this report is written it is very difficult to identify new fundamental issues of efficient algorithms in computer vision. However, lots of interesting aspects were discussed during the talks and in the discussion groups, and many participants established collaborations with people from other fields. Thus we believe this seminar served as an excellent basis to inspire and trigger novel developments in the design of efficient algorithms for global optimisation problems in computer vision.

Finally, it should be mentioned that there was a huge consensus among the participants that there should be a follow-up event to this seminar in the upcoming years. For the current seminar, there will be edited post-proceedings in the Springer LNCS series which gives all participants the opportunity to summarise results from the seminar, discuss open questions and present recent research in the field. The deadline for submission is end of March 2012.

2 Table of Contents

3 Overview of Talks

3.1 Simultaneous Convex Optimization of Regions and Region Parameters in Image Segmentation Models

Egil Bae (University of Bergen, NO)

Joint work of Egil Bae, Jing Yuan, Xue-Cheng Tai

This work develops a convex optimization framework for image segmentation models, where both the unknown regions and parameters describing each region are part of the optimization process. Convex relaxations and optimization algorithms are proposed, which produce results that are independent from the initializations and closely approximate global minima. We focus especially on problems where the data fitting term depends on the mean or median image intensity within each region. We also develop a convex relaxation for the piecewise constant Mumford-Shah model, where additionally the number of regions is unknown. The approach is based on optimizing a convex energy potential over functions defined over a space of one higher dimension than the image domain.

3.2 Multi-label Segmentation with Constraints on Geometry and Complexity

Yuri Boykov (Middlesex College – Ontario, CA)

Joint work of Andrew Delong, Lena Gorelick, Anton Osokin, Hossam Isack, Yuri Boykov

Existing segmentation methods rutinely apply mixture models to objects containing spatially distinct regions, each with a unique color/texture model. Since mixture models assume i.i.d. distribution of intensities, the spatial distribution of colors within complex objects is ignored, which significantly limits the robustness and the accuracy of segmentation results. Despite, mixture models are widely used in segmentation because they easily integrate into standard optimization techniques as simple unary potentials.

We discuss segmentation models corresponding to more general non-i.i.d. appearances and propose the corresponding optimization methods. Sub- labels are used for object parts with additional constraints needed to describe the distribution of parts. We focus on two types of generic constraints. The first type models geometric relations between the parts (e.g. "inclusion", "exclusion", etc. – useful in bio-medical applications due to anatomy). The second type models complexity of appearance (i.e. MDL principle – useful for both photo and medical imagery). These segmentation models allow either global optimization or sufficiently good approximations.

The talk is based on materials from four recent papers

- http://www.csd.uwo.ca/faculty/yuri/Abstracts/emmcvpr11-sabs.html
- http://www.csd.uwo.ca/faculty/yuri/Abstracts/ijcv10_lc-abs.html
- http://www.csd.uwo.ca/faculty/yuri/Abstracts/iccv11-abs.html
- http://www.csd.uwo.ca/faculty/yuri/Abstracts/iccv09-abs.html

3.3 Total Generalized Variation: A Convex Model for Piecewise Smooth Images

Kristian Bredies (Universität Graz, AT)

Joint work of Kristian Bredies, Martin Holler
Main reference K. Bredies, M. Holler, "Artifact-free JPEG decompression with total generalized variation," Int'l Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP), 2012.
A preprint is available as SFB-Report No. 2011-029 from the SpezialForschungsBereich F32.
URL http://math.uni-graz.at/mobis/publications/SFB-Report-2011-029.pdf

We introduce and study the total generalized variation functional (TGV) which constitutes a well-suited convex model for piecewise smooth images. It comprises exactly the functions of bounded variation but is, unlike total-variation based functionals, also aware of higher-order smoothness. The feasibility and the effect of regularization with TGV is demonstrated. In particular, we address the design of global optimization methods with TGV penalty. Furthermore, the framework is applied to the reconstruction of color images from inexact and incomplete data: We propose a TGV-based approach for the reduction of typical decompression artifacts in highly compressed JPEG images and show numerical results.

3.4 Completely Convex Formulation of the Chan-Vese Image Segmentation Model

Xavier Bresson (City University – Hong Kong, HK)

The active contours without edges model of Chan and Vese is a popular method for computing the segmentation of an image into two phases. The minimization problem is non-convex even when the optimal region constants are known a priori. In [Chan-Esedoglu-Nikolova 2006], authors provided a method to compute global minimizers by showing that solutions could be obtained from a convex relaxation.

In this talk, we propose a convex relaxation approach to solve the case in which both the segmentation and the optimal constants are unknown for two phases and multiple phases. In other words, we propose a convex relaxation of the popular K-means algorithm. Although the proposed relaxation technique is not guaranteed to find exact global minimizers of the original problem, our experiments show that our method computes tight approximations of the optimal solutions.

3.5 Efficient Video Analysis Tools towards Large Scale Unsupervised Learning

Thomas Brox (Universität Freiburg, DE)

Joint work of Thomas Brox, Narayanan Sundaram, Peter Ochs, Jitendra Malik
Main reference T. Brox, J. Malik, "Object segmentation by long term analysis of point trajectories," European Conference on Computer Vision (ECCV), 2010.
URL http://dx.doi.org/10.1007/978-3-642-15555-0_21

Learning tasks such as object detection require access to more training data than what is available today. One way to reduce the burden of manual image annotation is crowd sourcing; an alternative one is learning from videos. In particular, I will present video analysis tools that allow for long term analysis of videos by means of point trajectories. A clustering of these point trajectories enables automatic segmentation of moving objects, thereby providing annotation without any user interaction. Application of these tools to large amounts of data requires efficient algorithms that can make use of cheap parallel hardware, such as GPUs.

3.6 Continuous Limits of Discrete Perimeters

Antonin Chambolle (Ecole Polytechnique – Palaiseau, FR)

Main reference A. Chambolle, A. Giacomini, L. Lussardi, "Continuous limits of discrete perimeters," ESAIM: M2AN, 44(2), pp. 207–230, 2010.
URL http://dx.doi.org/10.1051/m2an/2009044

We have considered in this talk the continuous limit of various discrete approximations of the perimeter and/or total variation. After a discussion on how good can an isotropic approximation be, we have studied a class of discrete convex functionals which satisfy a (generalized) coarea formula. These correspond to approximations with discrete submodular functions, of not necessarily isotropic but translational invariant—homogeneous—perimeters. In these cases, indeed, perimeter minimization problems can be solved exactly with polynomial algorithms. We have described the variational limit, as the size of the pixels go to zero, of such perimeters, and shown in particular that they always correspond to a crystalline interfacial energy (which has an explicit form). A question which is still unclear is whether the limits of arbitrary submodular interactions form a larger set than the limits of approximations with pairwise interactions (which can be solved by max flow/min cut algorithms on a graph).

3.7 A Variational Approach for Exact Histogram Specification

Raymond Chan (Chinese University of Hong Kong, HK)

We focus on exact histogram specification when the input image is quantified. The goal is to transform this input image into an output image whose histogram is exactly the same as a prescribed one. In order to match the prescribed histogram, pixels with the same intensity level in the input image will have to be assigned to different intensity levels in the

output image. An approach to classify pixels with the same intensity value is to construct a strict ordering on all pixel values by using auxiliary attributes. Local average intensities and wavelet coefficients have been used by the past as the second attribute. However, these methods cannot enable strict-ordering without degrading the image.

In this paper, we propose a variational approach to establish an image preserving strict-ordering of the pixel values. We show that strict-ordering is achieved with probability one. Our method is image preserving in the sense that it reduces the quantization noise in the input quantified image. Numerical results show that our method gives better quality images than the existing methods.

3.8 Variational Methods for Image Reconstruction in Magnetic Resonance Imaging

Christian Clason (Universität Graz, AT)

Joint work of Christian Clason, Florian Knoll, Kristian Bredies
Main reference F. Knoll, C. Clason, K. Bredies, M. Uecker, R. Stollberger, "Parallel imaging with nonlinear reconstruction using variational penalties," Magnetic Resonance in Medicine, 67(1), pp. 34–41, 2012.
URL http://dx.doi.org/10.1002/mrm.22964

Acquisition speed is currently the main limiting factor for magnetic resonance imaging. One well established technique for scan time reduction acquires only a subset of the necessary data, but with multiple independent receivers, each of which has a different spatial profile which needs to be estimated. Standard reconstruction methods in current clinical practice are linear algebraic in nature and suffer from residual artifacts due to missing information and incorrectly estimated modulations. This talk demonstrates that a variational approach including penalties of total (generalized) variation type can significantly improve the reconstruction quality for accelerated image acquisition.

3.9 Geometrically Consistent Elastic Matching of 3D Shapes: A Linear Programming Solution

Daniel Cremers (TU München, DE)

Joint work of Thomas Windheuser, Ulrich Schlickewei, Frank R. Schmidt, and Daniel Cremers
Main reference T. Windheuser, U. Schlickewei, Frank R. Schmidt, D. Cremers, "Geometrically consistent elastic matching of 3D shapes: a linear programming solution," IEEE Int'l Conf. on Computer Vision (ICCV), 2011.
URL http://cvpr.in.tum.de/_media/spezial/bib/wssc_iccv11.pdf

We propose a method for computing a geometrically consistent and spatially dense matching between two 3D shapes. Rather than mapping points to points we consistently match infinitesimal surface patches in a manner which minimizes an elastic thin-shell energy, thereby minimizing the physical energy required to deform one shape into the other. We observe that any conceivable matching corresponds to a codimension-2 surface in the 4D product space spanned by the two shapes. Subsequently we show that a discrete formulation

of this problem is given by a binary linear program whose relaxed version can be solved efficiently in a globally optimal manner.

Experimental results demonstrate both qualitatively and quantitatively that the proposed LP relaxation allows to compute consistent and accurate matchings which reliably put into correspondence highly deformed and articulated 3D shapes consisting of thousands of triangles.

3.10 Convex Relaxations for Vectorial Multilabel Problems and the Total Curvature

Bastian Goldluecke (TU München, DE)

We discuss two (unrelated) convex relaxation techniques we recently introduced.

The first is for multilabel problems with a large number of vectorial labels, where we can drastically reduce the dimensionality of the problem by exploiting the special structure of the space. The second relaxation is for a novel continuous regularizer we call the Total Curvature of a function, which allows to use this notion of curvature as a replacement for total variation in image processing applications.

3.11 Are Extrapolation Methods Useful for Diffusion-based Image Analysis?

Sven Grewenig (Universität des Saarlandes, DE)

Diffusion methods are widely used in image processing and computer vision, e.g. for image denoising, interpolation, optic flow, and stereo reconstruction. There are two popular ways to implement them: Explicit finite difference schemes are simple but become inefficient due to severe time step size restrictions, while semi-implicit schemes are more efficient but require to solve large linear systems of equations. An efficient alternative is the recently proposed Fast Explicit Diffusion (FED) scheme that allows very large explicit time steps.

However, all three methods have first order accuracy in time. In this talk, we are going to consider more accurate time extrapolation methods based on semi-implicit and FED schemes. To this end, we present two stable FED extrapolation methods and analyse them with respect to their accuracy and efficiency for parabolic image enhancement as well as elliptic image interpolation problems.

3.12 Nonconvex TV^q-Models in Image Restoration: Analysis and a Trust-Region Regularization Based Superlinearly Convergent Solver

Michael Hintermueller (HU Berlin, DE)

Joint work of Michael Hintermueller, Tao Wu
Main reference M. Hintermüller, T. Wu, "Nonconvex TV^q-Models in Image Restoration: Analysis and a Trust-Region Regularization Based Superlinearly Convergent Solver," SFB-Report No. 2011-026, SpezialForschungsBereich F32, 2011.
URL http://math.uni-graz.at/mobis/publications/SFB-Report-2011-026.pdf

A nonconvex variational model is introduced which contains the l^q-norm, with $0 <= q <= 1$, of the gradient of the image to be reconstructed as the regularization term together with a least-squares type data fidelity term which may depend on a possibly spatially dependent weighting parameter. Hence, the regularization term in this functional is a nonconvex compromise between the minimization of the support of the reconstruction and the classical convex total variation model. In the discrete setting, existence of a minimizer is proven, a Newton-type solution algorithm is introduced and its global as well as locally superlinear convergence are established. The potential indefiniteness (or negative definiteness) of the Hessian of the objective during the iteration is handled by a trust region based regularization scheme. The performance of the new algorithm is studied by means of a series of numerical tests. For the associated infinite dimensional model an existence result based on the weakly lower semicontinuous envelope is given and its relation to the original problem is discussed.

3.13 Unsupervised Multiphase Segmentation and Its Regularity

Sung Ha Kang (Georgia Institute of Technology, US)

Variational models have been widely studied for imaging applications in recent years, and one of the most well known model is the Mumford-Shah functional was introduced in the late 1980s. In this talk, we discuss mainly two models. The first model is infinite perimeter segmentation model which is a variant of piecewise constant Mumford-Shah model (typically referred to as Chan-Vese model). It allows fine details and irregularities of the boundaries while denoising additive Gaussian noise. The second model is a multiphase segmentation model which automatically chooses a favorable number of phases as it segments the image. The model is discussed, connection is made with a regularized k-means (for 1-dimentional data clustering application), and new regularity analysis results are presented.

3.14 Metric Geometry In Action

Ron Kimmel (Technion – Haifa, IL)

I will talk about methods imported from metric geometry that help in comparing and matching shapes. Picking up the right metric for the task at hand plays an important role in shape analysis. I will review some tools for that goal.

3.15 Efficient Imaging Algorithms on Many-Core Platforms

Harald Koestler (Universität Erlangen-Nürnberg, DE)

Main reference H. Koestler, "A Multigrid Framework for Variational Approaches in Medical Image Processing and Computer Vision," Dr. Hut Verlag, Muenchen, 2008.

Today it is possible to achieve real-time performance for medium to large size imaging applications even on usual desktop computers or small clusters. For numerical algorithms like multigrid that work on structured grids GPUs show good speedup factors compared to current CPUs mainly due to their higher memory bandwidth. But also other imaging applications like sparse coding profit from GPUs.

Besides performance results for high dynamic range compression and image denoising we will also present some newer ideas for efficient image deblurring algorithms.

3.16 Maxflow-based Inference for Functions with High-order Terms

Vladimir Kolmogorov (IST Austria – Klosterneuburg, AT)

The maxflow technique is an important tool for minimizing functions with pairwise terms. It can compute a global minimum if all terms are submodular, and for general pairwise terms it can identify a part of an optimal solution by solving the roof duality relaxation. I will consider extensions of these techniques to higher-order terms. Previously proposed extensions converted the function to a sum of pairwise terms by introducing auxiliary variables. I will discuss a more direct approach that does not require such a conversion.

3.17 Volumetric Nonlinear Anisotropic Diffusion on GPUs

Arjan Kuijper (FhG IGD – Darmstadt, DE)

Joint work of Andreas Schwarzkopf, Thomas Kalbe, Chandrajit Bajaj, Arjan Kuijper, Michael Goesele
Main reference T. Kalbe, A. Schwarzkopf, M. Goesele, A. Kuijper, C. Bajaj, "Volumetric Nonlinear Anisotropic Diffusion on GPUs," Scale Space and Variational Methods in Computer Vision (SSVM), 2012.
URL http://dx.doi.org/10.1007/978-3-642-24785-9_6

We present an efficient implementation of volumetric nonlinear anisotropic image diffusion on modern programmable graphics processing units (GPUs). We avoid the computational bottleneck of a time consuming eigenvalue decomposition in R3.

Instead, we use a projection of the Hessian matrix along the surface normal onto the tangent plane of the local isodensity surface and solve for the remaining two tangent space eigenvectors. We derive closed formulas to achieve this resulting in efficient GPU code. We show that our most complex volumetric nonlinear anisotropic diffusion gains a speed up of more than 600 compared to a CPU solution.

3.18 Enforcing Topological Constraints in Random Field Image Segmentation

Christoph Lampert (IST Austria – Klosterneuburg, AT)

Joint work of Chao Chen, Daniel Freedman, Christoph H. Lampert
Main reference C. Chen, D. Freedman, C. H Lampert, "Enforcing topological constraints in random field image segmentation," IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2011.
URL http://dx.doi.org/10.1109/CVPR.2011.5995503

We introduce TopoCut, a new way to integrate knowledge about topological properties (TPs) into random field image segmentation model. Instead of including TPs as additional constraints during minimization of the energy function, we devise an efficient algorithm for modifying the unary potentials such that the resulting segmentation is guaranteed with the desired properties.

Our method is more flexible in the sense that it handles more topology constraints than previous methods, which were only able to enforce pairwise or global connectivity. In particular, our method is very fast, making it for the first time possible to enforce global topological properties in practical image segmentation tasks.

3.19 Image Restoration under Hessian Matrix-Norm Regularization

Stamatis Lefkimmiatis (EPFL – Lausanne, CH)

Joint work of Stamatis Lefkimmiatis; Aurelien Bourquard; Michael Unser
Main reference S. Lefkimmiatis, A. Bourquard, M. Unser, "Hessian-based norm regularization for image restoration with biomedical applications," IEEE Trans. on Image Processing, 2012, *to appear*.
URL http://dx.doi.org/10.1109/TIP.2011.2168232

We will discuss about non-quadratic Hessian-based regularization methods that can be effectively used for inverse imaging problems, in a variational framework. Motivated by the great success of the total-variation (TV) functional in the image-restoration setting, we will present an extension that also includes second-order differential operators. Specifically, we will focus on second-order functionals that involve matrix norms of the Hessian operator.

Their definition is based on an alternative interpretation of TV that relies on mixed norms of directional derivatives. We will show that the resulting regularizers retain some of the most favorable properties of TV, namely, convexity, homogeneity, rotation, and translation invariance, while dealing effectively with the staircase effect. We will further present an efficient minimization scheme for the corresponding objective functions. The employed algorithm is of the iterative-reweighted-least-squares (IRLS) type and results from a majorization-minimization approach. It also relies on a problem-specific preconditioned conjugate-gradient method (PCG) that makes the overall minimization scheme very attractive, since it can be effectively applied to large images in a reasonable computational time.

3.20 Optimality Bounds and Optimization for Image Partitioning

Jan Lellmann (Universität Heidelberg, DE)

Joint work of Jan Lellmann, Frank Lenzen, Christoph Schnörr
Main reference J. Lellmann, F. Lenzen, C. Schnörr, "Optimality bounds for a variational relaxation of the image partitioning problem," arXiv:1112.0974v1 [cs.CV]
URL http://arxiv.org/abs/1112.0974v1

Variational convex relaxations can be used to compute approximate minimizers of optimal partitioning and multiclass labeling problems on continuous domains. We present recent developments on proving a priori upper bounds for the objective of the relaxed problem, and on the numerical solution of such problems.

3.21 A Variational Approach for the Fusion of Exposure Bracketed Images

Stacey Levine (Duquesne University, US)

Joint work of Stacey Levine, Marcelo Bertalmio

We propose a variational approach for fusing a set of images taken with different exposure times so that optimal information is obtained from each one. The solution is a single image

whose details and edges are extracted from a short exposure time image (typically low contrast and noisy) and color information is extracted from a long exposure time image (often suffering from blur due to motion by the camera and/or the subject(s)). The approach is well posed and results compare favorably with the state of the art.

3.22 Total Variation Denoising Using Posterior Expectation

Cecile Louchet (Université d'Orléans, FR)

Total variation image denoising was originally described as a variational method, but it can be interpreted in a Bayesian framework as a Maximum A Posteriori estimate. This maximization aspect is partly responsible for the so-called staircasing effect, i.e. the outbreak of quasi-constant regions separated by sharp edges in the intensity map.

In this talk we propose to transpose this denoising method into an estimation based on the posterior expectation, in order to better account for the global properties of the posterior distribution. Theoretical and numerical results are presented, which demonstrate in particular that images denoised with the proposed scheme do not suffer from the staircasing effect. We then focus on the practical computation of TV-LSE, and propose an MCMC (Monte-Carlo Markov Chain) algorithm whose convergence is carefully analyzed. Both the model and the algorithm are flexible enough to be directly applied to other low-level image processing tasks, such as image deblurring and interpolation.

3.23 Hyperelasticity in Correspondence Problems

Jan Modersitzki (Universität zu Lübeck, DE)

Image registration is one of the challenging problems in image processing. Given two images taken for example at different times, from different devices or perspectives, the goal is to determine a reasonable transformation, such that a transformed version of one of the images is similar to the second one.

The problem is typically phrased in a variational setting for the wanted transformation field where meaningful regularization is a key issue. In this talk, we present an extension of the well-known linear elastic regularizer. Hyperelasticity is used and enables the modelling of large and non-linear transformations and ensures diffeomorphic mappings. The price to be paid is a non-convex but polyconvex objective function. We also Moreover, a stable and efficient numerical scheme is challenging.

In this talk, we present an introduction to hyperelastic image registration with applications from cardiac PET imaging. We introduce a new numerical framework.

This framework is based on the discretize then optimize paradigm and uses a sophisticated computation of the discrete analogues of the three invariants of the transformation tensor. We show several numerical examples that illustrate the potential of the hyperelastic regularizer as well as some application where hyperelastic regularization is mandatory.

3.24 Fast Regularization of Matrix-Valued Images

Guy Rosman (Technion – Haifa, IL)

Joint work of Guy Rosman, Yu Wang, Xue-Cheng Tai, Ron Kimmel, Alfred M. Bruckstein
Main reference G. Rosman, Y. Wang, X.-C. Tai, R. Kimmel, A. M. Bruckstein, "Fast regularization of matrix-valued images," UCLA CAM Report 11-87, December 2011.
URL ftp://ftp.math.ucla.edu/pub/camreport/cam11-87.pdf

Regularization of matrix-valued data is important in many fields, such as medical imaging, motion analysis and scene understanding, where accurate estimation of diffusion tensors or rigid motions is crucial for higher-level computer vision tasks. In this report we describe a novel method for efficient regularization of matrix group-valued images.

Using the augmented Lagrangian framework we separate the total-variation regularization of matrix-valued images into a regularization and projection steps, both of which are fast and parallelizable. Furthermore we extend our method to a high-order regularization scheme for matrix-valued functions.

We demonstrate the effectiveness of our method for denoising of several group-valued image types, with data in SO(n), SE(n), and SP D(n), and discuss its convergence properties.

3.25 Massively Parallel Multigrid for Image Processing

Ulrich Ruede (Universität Erlangen-Nürnberg, DE)

Multigrid is among the most efficient solvers for global problems, since the hierarchical system of coarser grids provide an elegant means to exchange information in the problem domain. Many variational image processing problems lead to elliptic differential equations that need this kind of data exchange. While the coarser grids are therefore essential for fast global solvers, it may become a problem for the parallelization. In the presentation we will demonstrate that a careful implementation can nevertheless show excellent scalability and provide powerful solvers also for very large image data sets.

3.26 Quadrature Rules, Discrepancies and Variational Dithering

Gabriele Steidl (TU Kaiserslautern, DE)

This paper deals with continuous-domain quantization which aims to create the illusion of a gray-value image by appropriately distributing black dots. For lack of notation we refer to the process as halftoning which is usually associated with the quantization on a discrete grid. Recently a framework for this task was proposed by minimizing an attraction-repulsion functional consisting of the difference of two continuous, convex functions. One of them describes attracting forces caused by the image gray values, the other one enforces repulsion between dots. In this paper, we generalize this approach by considering quadrature error functionals on reproducing kernel Hilbert spaces (RKHSs) with respect to the quadrature

nodes, where we ask for optimal distributions of these nodes. For special reproducing kernels these quadrature error functionals coincide with discrepancy functionals. It turns out that the attraction-repulsion functional appears for a special RKHS of functions on $\mathbb{R}^2$. Moreover, our more general framework enables us to consider optimal point distributions not only in $\mathbb{R}^2$ but also on the torus $\mathbb{T}^2$ and the sphere $\mathbb{S}^2$. For a large number of points the computation of such point distributions is a serious challenge and requires fast algorithms. To this end, we work in RKHSs of bandlimited functions on $\mathbb{T}^2$ and $\mathbb{S}^2$. Then, the quadrature error functional can be rewritten as a least squares functional. We propose a nonlinear conjugate gradient method to compute a minimizer of this functional and show that each iteration step can be computed in an efficient way by fast Fourier transforms at nonequispaced nodes on the torus and the sphere. Numerical examples demonstrate the good quantization results obtained by our method.

3.27 Generalized Ordering Constraints for Multilabel Optimization

Evgeny Strekalovskiy (TU München, DE)

Joint work of Evgeny Strekalovskiy, Daniel Cremers
Main reference E. Strekalovskiy, D. Cremers, "Generalized ordering constraints for multilabel optimization," IEEE Int'l Conf. on Computer Vision (ICCV), pp. 2619–2626, 2011.
URL http://dx.doi.org/10.1109/ICCV.2011.6126551

We propose a novel framework for imposing label ordering constraints in multilabel optimization. In particular, label jumps can be penalized differently depending on the jump direction. In contrast to the recently proposed MRF-based approaches, the proposed method arises from the viewpoint of spatially continuous optimization. It unifies and generalizes previous approaches to label ordering constraints:

Firstly, it provides a common solution to three different problems which are otherwise solved by three separate approaches. We provide an exact characterization of the penalization functions expressible with our approach.

Secondly, we show that it naturally extends to three and higher dimensions of the image domain. Thirdly, it allows novel applications, such as the convex shape prior. Despite this generality, our model is easily adjustable to various label layouts and is also easy to implement. On a number of experiments we show that it works quite well, producing solutions comparable and superior to those obtained with previous approaches.

3.28 Parallel Preconditioners for GPU-Multigrid Solvers

Robert Strzodka (MPI für Informatik – Saarbrücken, DE)

Joint work of Dominik Göddeke, Robert Strzodka
Main reference D. Göddeke, R. Strzodka, "Mixed precision GPU-multigrid solvers with strong smoothers," in Jack J. Dongarra, David A. Bader, and Jakub Kurzak (eds.): Scientific Computing with Multicore and Accelerators, pp. 131–147, CRC Press, 2010.

GPUs perform best on regular, independent work-loads while effective preconditioners ask for complex, serially coupled computations. The extreme cases of best hardware performance

on slowly converging numerical schemes or quickly converging schemes with slow serial execution are poor choices. The talk discusses how to find the middle ground by challenging the hardware with more complex data dependencies and at the same time relaxing purely serial dependencies with parallel variants. For structured matrices, we can now solve very ill-conditioned linear equation systems that were intractable with GPU hardware before.

References

1 D. Göddeke and R. Strzodka. Cyclic reduction tridiagonal solvers on GPUs applied to mixed precision multigrid. *IEEE Transactions on Parallel and Distributed Systems (TPDS), Special Issue: High Performance Computing with Accelerators*, 22(1):22-32, January 2011

3.29 Image Segmentation by Iteratively Optimization of Multiphase Multiple Piecewise Constant Model and Four-Color Relabeling

Wenbing Tao (Huazhong University of Science & Technology, CN)

An iteratively unsupervised image segmentation algorithm is developed, which is based on our proposed multiphase multiple piecewise constant (MMPC) model and its graph cuts optimization. The MMPC model use multiple constants to models each phase instead of one single constant used in Chan and Vese (CV) model and cartoon limit so that heterogeneous image object segmentation can be effectively deal with.

We show that the multiphase optimization problem based on our proposed model can be approximately solved by graph cuts methods. The Four-Color theorem is used to relabel the regions of image after every iteration, which makes it possible to represent and segment an arbitrary number of regions in image with only four phases. Therefore, the computational cost and memory usage are greatly reduced. The comparison with some typical unsupervised image segmentation methods using a large number of images from the Berkeley Segmentation Dataset demonstrates the proposed algorithm can effectively segment natural images with good performance and acceptable computational time.

3.30 Nonlocal Filters for Removing Multiplicative Noise

Tanja Teuber (TU Kaiserslautern, DE)

Joint work of Tanja Teuber, Annika Lang
Main reference T. Teuber, A. Lang, "Nonlocal filters for removing multiplicative noise," Scale Space and Variational Methods in Computer Vision (SSVM), 2012.

We present new nonlocal filters for removing multiplicative noise in images. The considered filters are deduced by weighted maximum likelihood estimation and the occurring weights are defined by a new similarity measure specially designed for comparing data corrupted by multiplicative noise. For the deduction of this measure we analyze a probabilistic measure recently proposed for general noise models by Deledalle et al. The properties of our new measure are examined theoretically as well as by numerical experiments. Moreover, different adaptations are proposed to further improve the results. Throughout the talk, our findings

are exemplified for multiplicative Gamma noise. Finally, restoration results are presented to demonstrate the good properties of our new filters.

3.31 Optimization for Pixel Labeling Problems With Structured Layout

Olga Veksler (University of Western Ontario, CA)

Joint work of Olga Veksler, Pedro Felzenszwalb, Xiaoqing Liu
Main reference P. F. Felzenszwalb, O. Veksler, "Tiered scene labeling with dynamic programming," IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2010.
URL http://dx.doi.org/10.1109/CVPR.2010.5540067

Pixel labeling problems are pervasive in computer vision research. In this talk, we discuss optimization approaches for labeling problems which have some structure imposed on the layout of the labels. In other words, the relationships between labels is not arbitrary but has a well defined spatial structure. We will describe two approaches for structured layout scenes.

The first approach is for a more restrictive type of scenes, for which we develop new graph-cut moves which we call order-preserving. The advantage of order preserving moves is that they act on all labels simultaneously, unlike the popular expansion algorithm, and, therefore, escape local minima more easily. The second approach is for a more general type of structured layout scenes and it is based on dynamic programming.

In the second case, the exact minimum can be found efficiently. This is very rare for a 2D labeling problem to have an efficient and global optimizer. For both approaches, our applications include geometric class labeling and segmentation with a shape prior.

3.32 3D Motion Analysis with Stereo Cameras

Andreas Wedel (Daimler AG, DE)

Joint work of Andreas Wedel, Clemens Rabe, Thomas Müller, Uwe Franke
Main reference C. Rabe, T. Müller, A. Wedel, U. Franke, "Dense, robust, and accurate motion field estimation from stereo image sequences in real-time," European Conf. on Computer Vision (ECCV), 2010.
URL http://dx.doi.org/10.1007/978-3-642-15561-1_42

We describe the estimation of the three dimensional motion field of the visible world from stereo image sequences. Our approach is build upon the ideas of "dense scene flow" and "6d vision". It combines dense variational optical flow estimation, including spatial regularization, with Kalman filtering for temporal smoothness and robustness. Using a GPU and an FPGA for the implementation yields a vision-system which is directly applicable in real-world scenarios, like automotive driver assistance systems or in the field of surveillance.

3.33 Discrete Theory and Numerical Analysis of Anisotropic Diffusion Interpolation

Joachim Weickert (Universität des Saarlandes, DE)

Joint work of Joachim Weickert, Sven Grewenig, Markus Mainberger, Martin Welk

Image compression with partial differential equations (PDEs) is based on the idea to store only a few pixels and to perform PDE-based inpainting inbetween. Experiments have shown that edge-enhancing anisotropic diffusion (EED) outperforms other PDEs for this task. However, it creates a number of numerical problems: There is no general energy functional, the process is highly anisotropic, dissipative artifacts may occur, and rotation invariance may be approximated badly. Since the theory of nonnegative matrices cannot be applied, no discrete maximum-minimum principle holds. Moreover, the process can be highly ill-conditioned, and linearisations of the nonlinear system of equations for the inpainting problem lead to unsymmetric system matrices. In my talk I will address these problems and show how some of them can be solved.

3.34 Variational Optical Flow

Manuel Werlberger (TU Graz, AT)

Joint work of Manuel Werlberger; Thomas Pock, Horst Bischof
Main reference M. Werlberger, T. Pock, H. Bischof, "Motion estimation with non-local total variation regularization," IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2010.
URL http://dx.doi.org/10.1109/CVPR.2010.5539945

The presented motion estimation algorithm is based on non-local total variation regularization which allows for integrating low level image segmentation in a unified variational framework. Based on Gestalt principles of grouping areas are combined together and the similarity of motion within these regions is enforced.

For a more robust data fidelity term we propose not to linearize the intensity function as for standard algorithms but to approximate the data term directly with a second order Taylor expansion. This allows for using arbitrary data terms.

3.35 Optimization on Spheres and Stiefel Manifolds, with Applications on Genus-0 Surfaces

Wotao Yin (Rice University – Houston, US)

Joint work of Xianfeng Gu; Lai, Lu Rongjie, Ming Lok; Zaiwen Wen, Wotao Yin

Minimization with orthogonality constraints ($X'X = I$) and/or spherical constraints ($||x||_2 = 1$) has wide applications in polynomial optimization, combinatorial optimization, eigenvalue problems, sparse PCA, p-harmonic flows, 1-bit compressive sensing, matrix rank minimization,

etc. These problems are non-convex, and it is numerically expensive to preserve the constraints while improving the objective.

We propose to use a Crank-Nicolson-like update scheme that preserves these constraints while performing descent on the manifold with curvilinear search. The update has a low per-iteration cost compared to those based on projections and geodesics. This talk describes the update formula, analyzes its complexity, and presents the full algorithm with the application of global conformal parameterization of genus-0 surfaces.

The optimization part is joint work with Zaiwen Wen. The conformal parameterization part is joint work with Rongjie Lai, Zaiwen Wen, Xianfeng Gu, and Lok Ming Lu. See http://optman.blogs.rice.edu for slides and preprints.

4 Working Groups

Based on proposals from the seminar participants three main working groups were established that met several times during the time of the seminar:

- Models and Priors
- Convex Methods
- Parallelization

Below we give detailed summaries of topics and outcomes of the different workings groups.

4.1 Models and Priors

The first working group addressed the fundamental question of designing models and priors for imaging problems.

In energy minimization approaches the cost functional often penalizes a certain norm on specific image features such as the image gradient or the residual between the noisy image and the unknown clean image. In the discussion of this working group, the following evolution of different norms was pointed out. Starting with the first work in image processing and computer vision in the eighties, mainly quadratic L^2 norms have been used in the cost functional. The main reason for using quadratic norms is based on the fact, that computing a minimizer of a quadratic cost functional amounts for solving systems of linear equations, which was a tractable problem at this time. With the development of new powerful optimization methods (mainly convex optimization algorithms) in the nineties, L^1 norm-type penalizers became very popular, mainly due to their robustness and edge-preserving properties. Later, it has been observed that so-called pseudo-norms (L^p norms for $1 < p < 1$) perform even better, but due to their non-convexity, these methods require much more sophisticated numerical optimization algorithms. Recently, the so-called L^0 norm became very popular in different branches of image and signal processing. Its basic idea is to find a very sparse representation of a signal in a suitable basis which allows for very accurate reconstructions from very few noisy measurements. In the end of this discussion, the question was raised what should come next, but no clear direction can be identified so far.

Another fundamental question that has been discussed in this working group was the question of higher-order methods, i.e. methods that penalize higher-order derivatives in images. While for almost three decades, methods penalizing first order derivatives are

dominating, it has been observed quite early, that higher-order methods provide a much stronger prior for imaging tasks. However, higher-order methods are much more demanding in terms of computational time and very often require the design of specialized numerical optimization algorithms. The consensus of this discussion was, that much more work has to be done into this direction to exploit the known potential of higher order methods.

Finally, the fundamental question of how to design or compute good image priors was discussed. On the one hand, researchers use "deterministic" image priors such as the total variation, which is known to favor piecewise constant images. This works well for images of this kind but clearly does not respect the diversity of natural images. Another approach, which is mainly used in image compression is to use local patch based discrete cosine transforms. This prior works much better for natural images but still does not yield satisfactorily results for many problems. On the other hand researchers try to "learn" good image priors. In recent work an algorithm is proposed, which simply fits a high-dimensional Gaussian mixture model to a large number of patches that have been extracted from an image database. it turns out that indeed a large portion of the structure that is learned coincides with oriented image derivatives or with oscillating patterns that can be found in the basis vectors of the discrete cosine transform. The working group concluded that there are fundamental relationships and that it is very important to investigate these methods.

4.2 Convex Methods

The second discussion group that attracted a large number of participants dealt with the design and mathematical analysis of convex formulations of computer vision problems.

So far, most of the global minimization algorithms for computer vision have been carried out through convexification of the minimization problem. One of the questions discussed is about how far this approach can research and be extended. The majority of the models proposed for image processing and computer vision need to minimize an energy functional that is non-linear and non-convex. Traditionally, algorithms often have problems with local minimizers and dependence on initial values. Some recent research results show that a number of these models can be solved through some convexified problems. The convex problems have global minimizers. These global minimizers are either global minimizers for the original non-conex problems or provide a good approximation. During the discussions, global minimization related to length regularization was discussed. Relationship between several methods recently proposed was reviewed. Potential methods that can overcome some problems related to length approximation were mentioned in the discussion. It is concluded that convexification approach can be used for a large number of imaging and vision problems. What we have discovered so far is just a few of them.

Another crucial issue for global minimization is to extend some of the recently proposed algorithms to higher order regularization techniques. Higher order regularizer are much more complicated. A few techniques have been proposed related to graphs. A number of talks in the workshop have discussed this issue. However, convexification of higher model related to "curvatures" and different norms of the Hessian of the minimization functional is still missing in the continuous setting. Researches and investigations into this direction need to be encouraged.

A third issue for convex global minimization is about efficiency and parallel computing. As imaging and vision problems are getting large with the size of data, it is urgent to improve the efficiency of the global minimization algorithms through different accelerations and

parallel implementation techniques. Some good acceleration techniques have been suggested in the literature. However, more research is needed on parallel acceleration techniques.

4.3 Parallelization

The third working group was discussing issues of parallelization, large scale computing and benchmarking.

Due to the strong increase in computational power and intrinsic parallelism, the use of graphics processor units (GPUs) became very popular to accelerate algorithms in the field of computer vision. In this context, papers on GPU optimisation typically report huge speed-ups of up to three orders of magnitude. In this discussion it became evident that in most cases the potential gain should actually be much smaller than reported in the literature – somewhere in the order between a factor two and five. This is due to the fact that most algorithms for global optimisation problems in computer vision are memory-limited such that speed ups are mainly related to an increase in memory bandwidth from CPU to GPU. This in turn means that many CPU algorithms are actually not efficiently implemented, either because researchers are not aware of standard CPU optimisation strategies (SIMD, cache optimisation), or because more time is spent on optimising the GPU code. As a conclusion for this first part, the members of the discussion group agreed that in general absolute run times should be considered instead of relative speed ups which may be highly misleading. Moreover, more research should be done on sequential algorithms before turning towards their parallel implementation.

With respect to large scale computing the discussion showed that the trend seems to go to cluster systems with millions of multi-core CPUs. This has three consequences: On the one hand, huge datasets for benchmarking are required that allow for a reasonable parallelization. Due to the increasing resolution of image acquisition devices such data sets are actually available (e.g. satellite images, video data, photo collections). On the other hand, large scale data sets require a sophisticated memory management based on adaptive or hierarchical strategies. This aspect is new to many researchers in computer vision that worked with small and medium scale problems so far. A third issue is the visualization of large scale data sets which becomes even more important due to the increasing gap between the size of data sets (up to a few petabyte) and the resolution of displays (almost constant at a few megapixels). Seemingly, working on compressed or even adaptively resampled data sets is another challenge that will become more and more important in computer vision. It is an open question how existing algorithms have to be modified such that they allow the processing of such condensed information. As a main conclusion of this second part the group members agreed that the algorithmic design pipeline described in the Dagstuhl proposal will get significantly more complex in the future. In the case of a potential follow up seminar, this should be addressed by inviting also people from the above mentioned fields.

Finally, the aspect of benchmarking the speed and the quality of efficient algorithms was discussed. Apart from the aforementioned problem that parallel implementations are typically compared to poorly optimised sequential schemes, a major issue in benchmarking is the fact that many researchers actually do a compromise between quality and speed when they evaluate the accuracy and runtime of their approach. Often the design of efficient algorithms is accompanied by the proposal of improved model components. In this case, some of the quality gain is sacrificed for improving the run time. Typically iterative schemes are stopped much too early and the results are far from convergence. In such cases it is not

clear how good the novel model actually performs and what part of the speed up refers to the underlying (potential parallel) implementation. It is evident that this problem is a general problem that is also relevant for sequential algorithms. In this context, the members of the discussion groups proposed that every research paper benchmarking efficient algorithms should provide at least one experiment where the true potential of an algorithm is made clear. This fully converged high quality result may then serve as a reference solution to judge the loss of quality when cutting down the runtime of the algorithm. Evidently, in the ideal case, no such loss should occur. However, one may observe that for some tasks approximate solutions obtained either by approximating the original problem or by stopping the algorithm too early are absolutely fine. In those cases it may even seem possible to obtain further speed ups by reducing the complexity of the original model. This, however, is another open question that came up during discussion and requires further research.

Participants

- Egil Bae
University of Bergen, NO
- Yuri Boykov
Middlesex College – Ontario, CA
- Kristian Bredies
Universität Graz, AT
- Xavier Bresson
City Univ. – Hong Kong, HK
- Thomas Brox
Universität Freiburg, DE
- Andrés Bruhn
Universität des Saarlandes, DE
- Antonin Chambolle
Ecole Polytechnique – Palaiseau, FR
- Raymond Chan
Chinese Univ. of Hong Kong, HK
- Christian Clason
Universität Graz, AT
- Daniel Cremers
TU München, DE
- Bastian Goldlücke
TU München, DE
- Sven Grewenig
Universität des Saarlandes, DE
- Michael Hintermueller
HU Berlin, DE
- Sung Ha Kang
Georgia Inst. of Technology, US
- Ron Kimmel
Technion – Haifa, IL
- Harald Köstler
Univ. Erlangen-Nürnberg, DE
- Vladimir Kolmogorov
IST Austria – Klosterneuburg, AT
- Arjan Kuijper
Fraunhofer Inst. – Darmstadt, DE
- Christoph Lampert
IST Austria – Klosterneuburg, AT
- Stamatis Lefkimmiatis
EPFL – Lausanne, CH
- Jan Lellmann
Universität Heidelberg, DE
- Stacey Levine
Duquesne University, US
- Cécile Louchet
Université d'Orléans, FR
- Arvid Lundervold
University of Bergen, NO
- Jan Modersitzki
Universität zu Lübeck, DE
- Thomas Pock
TU Graz, AT
- Guy Rosman
Technion – Haifa, IL
- Ulrich Rüde
Univ. Erlangen-Nürnberg, DE
- Gabriele Steidl
TU Kaiserslautern, DE
- Evgeny Strekalovskiy
TU München, DE
- Robert Strzodka
MPI für Informatik – Saarbrücken, DE
- Xue-Cheng Tai
University of Bergen, NO
- Wenbing Tao
Huazhong University of Science & Technology, CN
- Tanja Teuber
TU Kaiserslautern, DE
- Olga Veksler
Univ. of Western Ontario, CA
- Andreas Wedel
Daimler AG, DE
- Joachim Weickert
Universität des Saarlandes, DE
- Manuel Werlberger
TU Graz, AT
- Wotao Yin
Rice University – Houston, US

Report from Dagstuhl Seminar 11481

Models@run.time

Edited by

Uwe Aßmann[1], Nelly Bencomo[2], Betty H. C. Cheng[3], and Robert B. France[4]

1 **TU Dresden, DE,** `uwe.assmann@tu-dresden.de`
2 **INRIA Paris-Rocquencourt, FR,** `nelly@acm.org`
3 **Michigan State University, US,** `chengb@cse.msu.edu`
4 **Colorado State University, US,** `france@cs.colostate.edu`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 11481 "Models@run.time". Research on models@run.time seeks to extend the applicability of models and abstractions to the runtime environment, with the goal of providing effective technologies for managing the complexity of evolving software behaviour while it is executing. The Dagstuhl Seminar "Models@run.time" brought together a diverse set of researchers and practitioners with a broad range of expertise, including MDE, software architectures, reflection, self-adaptive systems, validation and verification, middleware, robotics and requirements engineering.

Seminar 27. November – 02. December, 2011 – www.dagstuhl.de/11481
1998 ACM Subject Classification D.2 Software Engineering, D.2.10 Design, H.1 Models and Principles
Keywords and phrases Self-adaptive Systems, Feedback Loop, Assurance, Uncertainty, Requirements, Optimization, Adaptation
Digital Object Identifier 10.4230/DagRep.1.11.91
Edited in cooperation with Sebastian Götz

1 Executive Summary

Uwe Aßmann
Nelly Bencomo
Betty H. C. Cheng
Robert B. France

To date, research on model-driven engineering (MDE) has mainly focused on the use of models during software development. This work has produced relatively mature techniques and tools that are currently being used in industry and academia to manage software complexity during development. Research on models@run.time seeks to extend the applicability of models and abstractions to the runtime environment, with the goal of providing effective technologies for managing the complexity of evolving software behaviour while it is executing.

As is the case for many software development models, a runtime model is often created to support reasoning. However, in contrast to development models, runtime models are used to reason about the operating environment and runtime behaviour for some purpose, for example, to determine an appropriate form of adaptation, and thus these models must capture abstractions of runtime phenomena. Different runtime dimensions need to be balanced when

adapting software at runtime, including efficient use of resources (time, memory, energy), context-dependencies (time, location, platform), and personalization concerns (quality-of-service specifications, profiles). The hypothesis can be stated as follows: Models@run.time that provide meta-information on these dimensions during execution enables the development of technologies that automate (1) runtime decision-making and (2) safe adaptation of runtime behaviour. Thus, we anticipate that this technology will play an integral role in the management of autonomic systems and self-adaptive systems.

The problems targeted by the models@run.time community are multi-faceted and thus tackling them requires expertise from a variety of research areas. The Dagstuhl Seminar models@run.time brought together a diverse set of researchers and practitioners with a broad range of expertise, including MDE, software architectures, reflection, self-adaptive systems, validation and verification, middleware, robotics and requirements engineering.

The following gives the objectives of the seminar and describes the extent to which they were met :

1. Objective: To identify and document the potential benefits of Models@run.time including benefits associated with their use in adaptive and autonomic systems. Extent Met: One of the working groups (Group 3: Uses and Purposes of M@RT) was charged with coming up with use cases that demonstrate the benefits associated with the use of models@run.time. The report produced by this group discusses the types of significant software systems that can benefit from use of models@run.time.
2. Objective: To reach a common understanding of the terminology and associated concepts that underpin the use of different models once a system is deployed. Extent Met: Each working group defined the terminology and the concepts used in the descriptions of their primary outcomes.
3. Objective: To identify a set of key research challenges that must be tackled to address the real-world problems posed by self-adaptive systems within the next five (5) years (the research roadmap). Extent Met: Each working group identified, discussed, and described the key challenges in its focus area. The identified challenges will be included in the roadmap we plan to publish.
4. Objective: To also identify associated technology transfer strategies to ensure that the research in this area has impact on industrial practice and associated methodologies and tool-sets. Extent Met: Group 2 (Runtime updating/adaptation mechanisms) focused on discussing and analysing the effectiveness of technologies that have been developed to support models@run.time.
5. Objective: To publish a collection of articles containing the roadmap, as well as papers from the participants. Extent Met: Plans have been put in motion for publishing peer-reviewed papers from participants in a LNCS State-of-the-Art Survey Volume on Models@run.time.

The seminar consisted of participant presentation and working group sessions. Monday and Tuesday morning were "speed dating" days, in which everyone (with the exception of some of the organizers, to better manage time) presented a 10-min introductory talk. The presenters covered a wide range of topics including adaptive cyber-physical systems, self-evolution, requirements-driven runtime adaptation, safe and trustworthy autonomous robots, adaptive and self-managing software, runtime variability and architectural reconfiguration at runtime. We also had two longer presentations by persons from industry (ERICSSON). The presenters were Joe Armstrong and Paer Emanuelsson. From Tuesday to Thursday participants worked in working groups that focused on particular aspects of models@run.time

research. At the start of each day the organizers summarized the activities that were to take place and the deliverables that were to be produced at the end of the day. At the end of each day, each group presented their deliverables.

On Wednesday evening we had a productive panel where the following questions were discussed: What are the compelling business models for models@run.time?, What are the killer applications? What are the obstacles to deployment of models@run.time systems?, What are key enabling technologies for models@run.time (e.g., standards or component models).

2 Table of Contents

Working Groups

3 Overview of Talks

3.1 Challenges in Designing models@runtime Systems and Our Related Research Activities

Mehmet Aksit (University of Twente, NL)

Runtime adaptable systems have become increasingly popular during the last decade, and there is a clear evidence that this trend will continue also in the future [1]. Designing effective and efficient models@runtime systems is far from trivial, since solutions to the design challenges demand combination of various techniques:

(i) there is a need for new languages and computational models that help in expressing such systems effectively. It has been shown that current object- oriented and aspect-oriented languages fall in short in this matter. Event- based languages offer considerable advantages [2,3]. Most runtime enforcement approaches (their objectives are similar to the ones of models@runtime) based on aspect-oriented languages support single base language (such as Java) and uni-process implementations and their expressive power is limited. We have developed various aspect-oriented (event-based) languages and systems (for example, Event-Composition Model and the EventReactor language) to overcome these limitations [4,5,6].

(ii) It is important to enforce the invariants of the models and the running systems and the conformance between the models and the systems. In [7], we show model checking of UML based models and conformance checking between the models and the running systems.

(iii) The control system(s) need to evaluate the running system from the perspective of various qualities (correctness, availability, memory usage, energy usage, performance etc.) and enforce optimal adaptation and control strategies. We have developed an architectural style and various techniques (for example based on multi-objective optimization) for this purpose [8,9,10].

(iv) We also believe that fuzzy-probabilistic techniques can be effective in dealing with uncertainty in such systems [11].

References

1 Aksit, M. and Choukair, Z. (2003) Dynamic Adaptive and Reconfigurable Systems Overview and Prospective Vision. In: Workshop on Distributed Auto-adaptive Reconfigurable Systems (DARES) – International Conference on Distributed Computing Systems (ICDCS), 23rd International Conference on Distributed Computing Systems Workshops (ICDCS'03 Workshops), 19-22 May 2003, Rhode Island, Providence, USA. pp. 84-89. IEEE Computer Society. ISBN 0-7695-1921-0

2 Malakuti Khah Olun Abadi, S. (2011) Event Composition Model: Achieving Naturalness in Runtime Enforcement. PhD thesis, University of Twente. CTIT Ph.D. thesis series no. 11-205 ISBN 978-90-365-3246-4

3 Bockisch, C.M. and Malakuti Khah Olun Abadi, S. and Aksit, M. and Katz, S. (2011) Making aspects natural: events and composition. In: Tenth International Conference on Aspect-Oriented Software Development, AOSD 2011, 21-25 Mar 2011, Porto de Galinhas, Brazil. pp. 285-299. ACM. ISBN 978- 1-4503-0605-8

4 Malakuti Khah Olun Abadi, S. and Aksit, M. and Bockisch, C.M. (2011) Runtime Verification in Distributed Computing. Journal of Convergence, 2 (1). pp. 1-10. ISSN 2093-7741

5 Malakuti Khah Olun Abadi, S. and Aksit, M. and Bockisch, C.M. (2011) Distribution-Transparency in Runtime Verification. In: Ninth IEEE International Symposium on Parallel

and Distributed Processing with Applications Workshops, ISPAW 2011, 26-28 May 2011, Busan, Korea. pp. 328-335. IEEE Communications Society. ISBN 978-1-4577-0524-3

6 Malakuti Khah Olun Abadi, S. and Bockisch, C.M. and Aksit, M. (2009) Applying the Composition Filter Model for Runtime Verification of Multiple- Language Software. In: The 20th annual International Symposium on Software Reliability Engineering,ISSRE 2009, 16-19 Nov 2009, Mysore, India. pp. 31-40. IEEE Computer Society. ISBN 978-0-7695-3878-5

7 Ciraci, S. and Malakuti Khah Olun Abadi, S. and Katz, S. and Aksit, M. (2010) Checking the Correspondence Between UML models and Implementation. In: 1st International Conference on Runtime Verification, 1-4 Nov 2010, Malta. pp. 198-213. Lecture Notes in Computer Science 6418. Springer Verlag. ISSN 0302-9743 ISBN 978-3-642-16611-2

8 Sozer, Tekinerdogan and Aksit. Optimizing Decomposition of Software Architecture for Local Recovery" for publication in Software Quality Journal.

9 de Roo, A.J. and Sözer, H. and Aksit, M. (2011) Runtime Verification of Domain-Specific Models of Physical Characteristics in Control Software. In: Fifth IEEE International Conference on Secure Software Integration and Reliability Improvement, SSIRI 2011, 27-29 Jun 2011, Jeju Island, Korea. pp. 41-50. IEEE Computer Society. ISBN 978-0-7695-4453-3

10 de Roo, A.J. and Sözer, H. and Aksit, M. (2009) An Architectural Style for Optimizing System Qualities in Adaptive Embedded Systems using Multi- Objective Optimization. In: Proceedings of the 8th Working IEEE/IFIP Conference on Software Architecture, 14-17 Sep 2009, Cambridge, UK. pp. 349-352. IEEE Computer Society. ISBN 978-1-4244-4985-9

11 Noppen, J.A.R. and van den Broek, P.M. and Aksit, M. (2008) Software development with imperfect information. Soft Computing – A Fusion of Foundations, Methodologies and Applications, 12 (1). pp. 3-28. ISSN 1432-7643

3.2 Programming systems that never stop

Joe Armstrong (ERICSSON – Stockholm, SE)

How can we program systems that never stop? Erlang (a programming language) and more specifically OTP (a set of libraries) was designed specifically for building fault-tolerant systems. The systems we have built in Erlang have been running for many years and during this time have been upgraded many times without significant loss of service. In this context "many years" means of the order of ten years and "many times" means hundreds to thousands of times. These are large systems servicing hundred to millions of simultaneous sessions. I'll talk about the architectures of such systems and how they are designed for dynamic code and service upgrade and how this works in practice.

3.3 Multi-quality Dynamic Auto-Tuning for Cyber-Physical Systems

Uwe Aßmann (TU Dresden, DE)

Cyber-physical systems of the future are safety-critical. They need verification of qualities but also need to be highly adaptive.

This requires completely new software architectures, based on dynamic multi-quality auto-tuning. Auto-tuning is an optimization method from high-performance computing which adapts an algorithm towards system and context parameters.

For cyber-physical systems, we suggest to extend auto-tuning to multiple qualities, based on quality contracts [1].

References

1 S. Götz, C. Wilke, M. Schmidt, S. Cech, and U. Aßmann. Towards Energy Auto-Tuning. In Proceedings of First Annual International Conference on Green Information Technology (GREEN IT), 2010.

3.4 Claim Monitoring for Tackling Uncertainty in Adaptive Systems

Nelly Bencomo (INRIA, FR)

Joint work of Bencomo, Nelly; Sawyer Pete; Welsh Kris

There is an increasing need for software systems that are able to adapt dynamically to changes in their environment.

However, a challenging characteristic of self-adaptive systems is that of uncertainty; a full understanding of all the environmental contexts they will encounter at runtime may be unobtainable at design time. Thus assumptions may have to be taken that risk being wrong, and this may lead to problems at runtime. We have developed REAssuRE, which uses the concept of claims to explicitly represent such assumptions in goal models of the system. We define a semantics for claims in terms of their impact on how alternative goal operationalisation strategies satisfy the system's non-functional requirements or soft goals. The implementation of REAssuRE includes automatic claim value propagation and goal model evaluation, using in-memory representations of the goal models and associated claims. We have demonstrated how claims can be monitored to verify claims at run time, and how falsified claims can trigger principled adaptation. We have evaluated REAssuRE using an adaptive flood warning system.

References

1 Kristopher Welsh, Pete Sawyer, Nelly Bencomo. *Towards Requirements Aware Systems: Run-time Resolution of Design- time Assumptions*. E26th IEEE/ACM International Conference On Automated Software Engineering, ASE 2011, Kansas, USA, 2011

3.5 Models@RT to support Interoperability

Amel Bennaceur (INRIA, FR)

Interoperability is a fundamental challenge for today's extreme distributed systems.Indeed, the high-level of heterogeneity in both the application layer and the underlying middleware and the conflicting assumptions that each system makes about its running environment

hinder the successful interoperation of independently-developed systems. Many solutions that aggregate the disparate systems in a non-intrusive way have been proposed.

These solutions use intermediary software entities, called mediators, to interconnect systems despite disparities in their data and/or interaction models by performing the necessary coordination and translations while keeping them loosely-coupled. Creating mediators requires a substantial development effort and a thorough knowledge of the application-domain, which is best understood by domain experts. Moreover, the increasing complexity of today's distributed systems, sometimes referred to as Systems of Systems, makes it almost impossible to develop 'correct' mediators manually. Therefore, formal approaches are used to automatically synthesize mediators. The notion of mediator is further realized using emergent middleware.

In this context, run-time models are used to capture meta-information about the heterogeneous systems including their interfaces and associated behaviour.

This is supplemented by ontological information to enable semantic interoperability in given application domains. We examine the nature of such run-time models coupled with consideration of the supportive synthesis algorithms that use these models to generate the appropriate mediators in order ensure interoperability in highly-heterogeneous environments.

This work takes place in the CONNECT project, which also examines how the models are derived, and how the system can adapt to underlying changes in context or issues related to the performance or behaviour of the system.

3.6 Self-Evolution under the Hood

Walter Cazzola (Università di Milano, IT)

No system escapes from the need of evolving either to fix bugs or to add new features. System evolution becomes particularly a problem in case the system cannot be stopped. The approach to solve this problem that is steadily gaining ground in the last few years is to grant reflective capability to the system that will introspect on its own model to plan its evolution when an update is necessary and drive the changes accordingly on its model and code.

Such an approach requires to fill the existing gap between the code and the model (often outdated and too loose) and to have an adequate support by the underlying framework (either the operating system or the virtual machine) [1]. The former is typically addressed by model driven engineering or reverse engineering techniques; the solution to the latter instead relies always on ad hoc solutions since the support to dynamic adaptation is quite limited.

To this respect, we developed JavAdaptor [3], a DSU approach which allows us to flexibly update running systems in an unanticipated manner.

JavAdaptor is a framework built on Java to deeply support dynamic evolution of Java applications; it permits to add/remove fields/methods (schema changes) and code adaptation with only a minimal lose of performance and without stopping a running application and above all the running application does not need to be pre-prepared to be updated by JavAdaptor (no special hooks are necessary in the code).

JavAdaptor, as it is, represents the perfect back-end for self-evolution when cooperates with an evolutionary planner that feeds it with the changes to apply to the running system.

Apart from presenting JavAdaptor and its characteristics, in this contribution, we are going to discuss its integration into a reflective architecture (a previous work from us named RAMSES [2]) to enable self-evolution driven by changes on the application model.

References

1 Walter Cazzola. Cogito, Ergo Muto! In *Proceedings of the Workshop on Self-Organizing Architecture (SOAR'09)*, pages 1–7, Cambridge, United Kingdom, September 2009.

2 Walter Cazzola, Ahmed Ghoneim, and Gunter Saake. Software Evolution through Dynamic Adaptation of Its OO Design. In *Objects, Agents and Features: Structuring Mechanisms for Contemporary Software*, Lecture Notes in Computer Science 2975, pages 69–84. Springer, July 2004.

3 Mario Pukall, Christian Kästner, Walter Cazzola, Sebastian Götz, Alexander Grebhahn, Reimar Schöter, and Gunter Saake. JavAdaptor — Flexible Runtime Updates of Java Applications. *Software—Practice and Experience*, 2012.

3.7 Challenges for Models@Runtime

Amit K. Chopra (University of Trento – Povo, IT)

My interest in models@runtime stems from my work with Tropos-like requirements models. These principal abstractions of these models are actors and their goals and commitments. I imagine them all to be runtime artefacts: actors would act toward satisfying their goals for which they would have to potentially enter into commitments with other actors. In this sense, I see an actor's goal model as serving essentially as the program for the actor. Although many challenges need to be addressed for this vision to become reality, it illustrates my take on models@runtime: introduce a new layer of abstraction (the model) with its own interpreter. The function of the interpreter is to translate model-level concepts into the appropriate operational ones. For example, goals may be operationalised into procedures, constraints, and so on. The technical challenges relate to understanding the concerns to be addressed in high-level models; judiciously selecting and formalizing the concepts required so as to not end up with a concept soup; and understanding the assumptions that can be made of the operational layer and the nature of the API with which its services may be invoked. The biggest challenge though remains in making sure that we are engineering systems to user requirements, no matter how high-level the abstractions we employ.

3.8 Urban-Scale, Dynamically Adaptable Systems

Siobhan Clarke (Trinity College – Dublin, IE)

Joint work of Clarke, Siobhan; Popescu, Razvan; Staikopoulos, Athanasios; Brennan, Shane; Fritsch, Serena; Groba, Christin; O'Toole, Eamonn; Hannon, Paula;
URL http://www.dsg.scss.tcd.ie/

My research interests are in design and programming models for dynamically adaptable systems. Systems of interest include large-scale, multi-layer systems, and distributed embedded

systems, including wireless sensor networks. In particular, I am investigating systems to support smart and sustainable cities, where resources, such as energy, transport and water, need to be optimised, and knowledge available from sensors and devices in the environment can be harnessed to support citizens' daily lives. Models at runtime will provide the view of the system and its environment that will enable analysis and triggering of appropriate adaptations to ensure the required optimisations occur in a timely manner. Research challenges include how to extract and filter the appropriate runtime data from the large scope of information available, and reason over that data in manner timely enough to effect appropriate adaptations.

3.9 A (Personal) Historical Perspective on Models@Run.time

Fabio M. Costa (Universidade Federal de Goiás, BR)

Joint work of Costa, Fabio M.; Blair, Gordon S.; Clarke, Peter J.

I started looking into the use of models at runtime when I was doing research on adaptive middleware for my PhD thesis at Lancaster University back in the year 2000. Reflection was a nice way to do adaptive middleware, but despite the use of meta-object protocols, it lacked a good approach to structure the self-representation of the base-level system. Then we came across with work from Kerry Raymond's group at DSTC, Australia, which proposed the concept of meta-information management, basically an approach for maintaining MOF-like repositories of meta-data (actually structured in the form of models) that could be accessed at runtime in a distributed environment [1]. To me, that was the seed to start thinking about ways to effectively use middleware models at runtime, and in the year 2000 we published a paper on the combination of reflection and meta-information management for adaptive middleware [2]. In that paper we proposed the design time use of models to generate middleware configurations, and, at runtime, the use of these same models as the causally connected self-representation of the middleware components that was maintained by the reflective meta-objects for the purposes of dynamic adaptation. This approach was realized as part of a middleware architecture (and a series of prototypes) called Meta-ORB, which also allowed the update of the runtime model and the creation of new model elements as a result of reflection. Now, after six successful editions of the Models@Run.time workshop, which built a very active community around the theme, and after this very interesting Dagstuhl seminar, it is good to see how the work of this great community has contributed to the development of the field to a point where it is gaining momentum and is starting to have an impact on the practice of building, maintaining and evolving software in many different application domains.

From the perspective of our research, some challenges that lie ahead include finding ways to deal with the complexity of models and developing efficient and safe (i.e., consistency-preserving) techniques to manipulate the models while the system is running, as well as to maintain the causal connection link without interfering with the performance of the system. The need to support user-centric models (models developed by non-technical users) and the requirement to manage distributed models are important additional challenges.

Our current research is looking into new approaches to use models at the core of adaptive/reflective and autonomic middleware technologies. We are particularly interested in the use of models to simulate and verify dynamic adaptations before they are actually

applied to the system. We are also looking into the development of a generic framework for model-driven middleware, using concepts such as layered architectures with models at different levels of abstraction, separation of behavioural and structural models (i.e., the model of the middleware system vs. the model run by the system), direct execution of (structural) models via interpretation, and the realization of functionality specified in user models via the dynamic selection and combination of existing underlying service providers. This generic framework is defined at the meta-model level and can be instantiated to realize the approach in different application domains. The approach originated from existing work on the Communication Virtual Machine middleware [3], and we are now working towards applying it to the domains of smart micro-grids and ubiquitous computing environments.

References

1 S. Crawley, S. Davis, J. Indulska, S. McBride, K. Raymond. 1997. Meta Information Management. In Proceedings 2 nd IFIP International Conference on Formal Methods for Open Object-based Distributed Systems (FMOODS'97).

2 Fabio M. Costa and Gordon S. Blair. 2000. Integrating Meta-Information Management and Reflection in Middleware. In Proceedings of the International Symposium on Distributed Objects and Applications (DOA '00). IEEE Computer Society, Washington, DC, USA, 133-.

3 Yali Wu, Andrew A. Allen, Frank Hernandez, Robert B. France, and Peter J. Clarke. A Domain-Specific Modeling Approach to Realizing User-Centric Communication. Journal of Software Practice and Experience (SP&E). Accepted February 2011.

3.10 Models at RunTime for the Certification of Autonomous Assistive Robots

Kerstin I. Eder (University of Bristol, GB)

Human-assistive robots are machines designed to improve the quality of our lives by helping us to achieve tasks, e.g. a personal care robot helps a patient in a hospital during recovery. Such robots perform physical tasks within the personal space of a human, including shared manipulation of objects and even direct contact. While the actions a robot performs are largely the same, every single execution will be slightly different in detail from the last, e.g. a nursing assistant serving lunch provides differing degrees of support during a patient's recovery. This requires robots to adapt their behaviour to different situations. To be genuinely useful, some robots may need to be powerful and therefore are potentially dangerous.

The development of human-assistive robots introduces new ethical, legal and societal issues. One fundamental concern is whether human-assistive robots can be trusted by humans. Essential components of trustworthiness are *usefulness* and *safety*; both have to be demonstrated for humans to gain confidence in the trustworthiness of such robots and certainly before such robots pass product certification. *How this can be done is currently an open research question.*

At the Bristol Robotics Lab I am conducting research to understand the verification and certification needs arising out of the latest developments in Autonomous Assistive Robots. My aim is to develop design and verification techniques as well as overall methodologies

leading to certification to push the state of the art in this area. My first projects address this issue in the context of Human-Robot Interaction.

Design Verification is a process used to demonstrate the correctness of a design w.r.t. the requirements and specification. Techniques used for design verification rely upon the requirements, specification and design to fully define the functional behaviour of the system at the time of verification. Human-assistive robots, however, are designed to adapt their behaviour. There is typically no complete description of the entire range of possible behavioural adaptations, which renders traditional verification techniques useless in this context. Even if the full set of behavioural descriptions was available, it would be so large that traditional design-time verification techniques would quickly reach their limits. Fig. 1 shows *Adaptability vs. Verifiability* ranging from a system that is fully adaptive but has too many behaviours making verification no longer tractable (on the left) to a system that is easily verifiable but not adaptive.

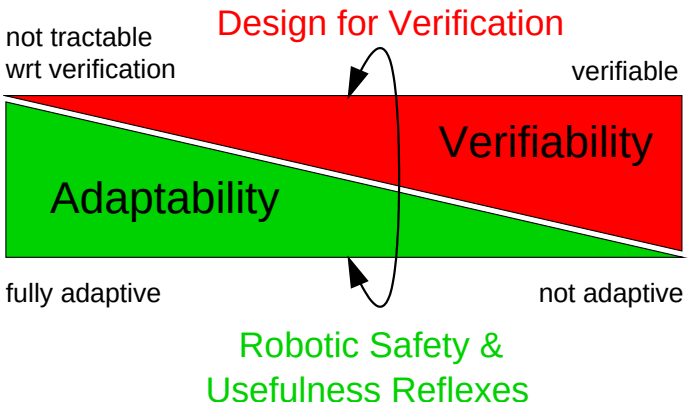

Figure 1 Adaptability vs. Verifiability

My research involves new techniques based on models at runtime to enable the verification of behavioural adaptations of human-assistive robots at runtime.

These models capture a consistent state of the agent initially (pre-verified warm start). Consistency is maintained at runtime. Based on the model the robot autonomously makes decisions. This enables the robot to dynamically adapt its behaviour to the situation such that it fulfils its requirements at all times.

The model contains a high-level description of the requirements and constraints that govern the functional (and non-functional) behaviour of the robot. It may be layered to reflect the robot's state, the state of the environment of the robot including the state of the humans/agents interacting with the robot, etc.

The key challenges for my research are currently:

- Establishing a set of high-level requirements that serve as invariant properties to be preserved during online behavioural adaptations, i.e. a set of invariant properties that form a "virtual cage"
- Online verification and validation (V&V) need to be tackled before certification: Can we design for online V&V? Can we build/adapt "correct-by-construction"?
- Finding a good compromise between the accuracy of the model to be a useful base for decision making and behavioural adaptation and computational complexity in terms of reasoning/processing based on the model.

3.11 Flexible Telecom systems that never stop

Paer Emanuelsson (Ericsson AB – Linköping, SE)

Telecom systems are large distributed systems with complex functionality that handle millions of users with real time performance. These systems evolve a lot during decades of years in operation. A system may only be unavailable for some minutes per year which makes the usual update procedures that stop the system impossible to use. Updates have to be made while the system is running.

We will briefly describe some different techniques used for achieving flexibility.

1. Parameters
2. Software update during runtime
   a. patches, data cannot be updated
   b. updates including data changes

1. Parameters

When it is known that there is need for variability a parameter is introduced.

By using parameters we avoid having lots of versions optimized for different markets and operators, which could lead to a maintenance nightmare. The biggest drawback with parameters is that changes have to be foreseen.

2a. Patches

Every software unit has a patch area, into which the patch can be loaded. The new software is activated by placing a jump to the new code.

2b. Updates including data changes

Software is run on two processors, and the second is a hot standby. While the first processor handles current traffic the new software is loaded onto the second processor. Runtime data is then transferred from the first processor to the second, which is specified in a data transformation language. This allows for arbitrary software changes. The data transformations necessary when data structures are changed can however be quite complex. The software doing the change from old to new version has to be designed for each change and the effort for specifying and performing the change can be larger than the development itself. It is also possible to have the new and old versions running simultaneously and transfer traffic only when needed – that is when old traffic does not end by itself after a period of waiting. Another way of handling data, which would allow for complex dependencies between different parts is to use a distributed real time database. Then transaction protected transformations can be performed.

3.12 Models@RT to support Interoperability

Nikolaos Georgantas (INRIA Le Chesnay, FR)

Interoperability is a fundamental challenge for today's extreme distributed systems. Indeed, the high-level of heterogeneity in both the application layer and the underlying middleware and the conflicting assumptions that each system makes about its running environment

hinder the successful interoperation of independently-developed systems. Many solutions that aggregate the disparate systems in a non-intrusive way have been proposed.

These solutions use intermediary software entities, called mediators, to interconnect systems despite disparities in their data and/or interaction models by performing the necessary coordination and translations while keeping them loosely-coupled. Creating mediators requires a substantial development effort and a thorough knowledge of the application-domain, which is best understood by domain experts. Moreover, the increasing complexity of today's distributed systems, sometimes referred to as Systems of Systems, makes it almost impossible to develop 'correct' mediators manually. Therefore, formal approaches are used to automatically synthesize mediators. The notion of mediator is further realized using emergent middleware.

In this context, run-time models are used to capture meta-information about the heterogeneous systems including their interfaces and associated behaviour. This is supplemented by ontological information to enable semantic interoperability in given application domains. We examine the nature of such run-time models coupled with consideration of the supportive synthesis algorithms that use these models to generate the appropriate mediators in order to ensure interoperability in highly-heterogeneous environments.

This work takes place in the CONNECT project, which also examines how the models are derived, and how the system can adapt to underlying changes in context or issues related to the performance or behaviour of the system.

3.13 Models at Runtime for Adaptive and Self-managing Software

Holger Giese (Hasso-Plattner-Institut – Potsdam, DE)

Software systems have to be continuously adapted to their changing requirements or environments. This is typically done by maintenance, while self-aware, context-aware, ultra-large-scale, or mission-critical software systems often have to be adapted during operation (self-adaptive software). Model-driven engineering and models at runtime play crucial roles in supporting maintenance, enabling self-adaptation, and integrating both, maintenance and self-adaptation.

In our talk, we outline our perspective on "models at runtime" as any models used on-line, i.e., internal to the running system, to represent running software, to represent the software's environment, or to manipulate or analyse any of the former two. Moreover, we sketch the benefits of models at runtime for maintenance, self-adaptation, and for an integrated setting of maintenance and self-adaptation. Considering different kinds of runtime models required for self-adaptation and self-adaptation activities as model operations, we discuss the role of a megamodel for specifying adaptation loops for self-adaptive software and operationalising the treatment of runtime models as means for monitoring, analysis, planning, and executing planned changes.

3.14 Multi-Quality Auto-Tuning with Contract Negotiation

Sebastian Götz (TU Dresden, DE)

Main reference Sebastian Götz, Claas Wilke, Sebastian Cech and Uwe Aßmann. Architecture and Mechanism for Energy Auto-Tuning. To appear in: Sustainable Green Computing. Practices, Methodologies and Technologies. IGI Global. 2012

Building Software Architectures for self-optimizing, quality-aware systems, requires to explicitly express the non-functional concerns and how in particular they interleave using the notion of models at runtime. In our Multi-Quality Auto-Tuning approach we distinguish structure and variant models. The first are used to describe the types of software, hardware and users that constitute the system. The second are used to describe the actual state of the system and to prescribe target states of the system. To make non-functional concerns explicit we use contracts, which describe the relation between different elements of the system in terms of their non-functional requirements as well as provisions. All these models are and have to be used at runtime to reason about the system and thereby to identify the most efficient configuration.

The process of reasoning faces a trade-off between response time and intelligence.

The major challenge from my point of view is finding the right abstraction-level for the models and fast, intelligent reasoning/optimization techniques.

3.15 Runtime Monitoring of Java Bytecode with UML and OCL Models

Martin Gogolla (Universität Bremen, DE)

Joint work of Hamann, Lars; Gogolla, Martin; Kuhlmann, Mirco
Main reference L. Hamann, M. Gogolla, M. Kuhlmann, "OCL-based Runtime Monitoring of JVM hosted Applications," in Proc. 11th OCL Workshop (2011), ECEASST Vol. 44, 20 pages.
URL http://journal.ub.tu-berlin.de/eceasst/article/view/623/677

Implementations of object-oriented software usually contain a lot of technical classes. Thus, the central parts of an application, e.g., the business rules, may be hidden among peripheral functionality like user-interface classes or classes managing persistence. Our approach makes use of modern virtual machines and allows the developer to profile an application in order to achieve an abstract monitoring and quality assurance of central application components during runtime. We represent virtual machine bytecode in form of a so-called platform-aligned model (PAM) comprising OCL invariants and pre- and postconditions. We show a prototype implementation as an extension of the UML and OCL tool USE.

3.16 Models@run.time for the proactive detection of QoS Problems

Lars Grunske (TU Kaiserslautern, DE)

Models@run.time are currently being used to evaluate quality demands and are especially suitable for performance, reliability, safety, and availability properties [2]. Our research is concerned with providing dedicated monitoring approaches to align the models with the running system [1,3]. Complemented with time-series analysis and change point detection techniques, these monitoring approaches allow for dynamic detection and proactive resolution of quality problems in modern software systems.

References

1 A. Amin, A. Colman, and L. Grunske. Using Automated Control Charts for the Runtime Evaluation of QoS Attributes. In *Proceedings of the 13ht IEEE International High Assurance Systems Engineering Symposium*. IEEE Computer Society, 2011.
2 R. Calinescu, L. Grunske, M. Z. Kwiatkowska, R. Mirandola, and G. Tamburrelli. Dynamic QoS management and optimization in service-based systems. *IEEE Trans. Software Eng.*, 37(3):387–409, 2011.
3 L. Grunske. An effective sequential statistical test for probabilistic monitoring. *Information and Software Technology*, 53(3):190 – 199, 2011.

3.17 SM@RT: A Model Diven Framework for Runtime Software Architecture

Gang Huang (Peking University, CN)

Runtime software architectures (RSA) are architecture-level, dynamic representations of running software systems, which help monitor and adapt the systems at a high abstraction level. The key issue to support RSA is to maintain the causal connection between architecture and system, ensuring that the architecture represents the current system, and the modifications on architecture cause proper system changes.

We developed a model-driven framework, called SM@RT (supporting models at runtime), to construct and maintain RSA in an automated way. SM@RT does four contributions to RSA: First, SM@RT gives a formal definition of RSA with a set of meta models and their relationships; Second, SM@RT automatically generates all codes implementing the causal connection between the system and RSA after developers define the software architecture model they preferred, the management model of the target runtime system, and the causal connection between them using the above meta models; Thirdly, SM@RT provides an incremental bidirectional transformation- based engine for synchronizing multiple RSA models; Fourth, SM@RT provides QVT as a default architecture manipulation language for reading, writing and analysing the RSA in a programmatic way.

SM@RT is compliant with MOF and QVT, the dominant standards of modelling technology, and then almost all artefacts produced with SM@RT are reusable and integrable, which reduce much engineering work in practice.

SM@RT has been experimented on diverse systems, from JEE enterprise systems (ECPerf, JPS and Rubis on JOnAS, JBoss and PKUAS), Java desktop systems (Eclipse GUI and XML Parser), to mobile systems (Android and PLASTIC/Window Mobile).

The future work of SM@RT will focus on seeking killer applications of models at runtime, in which the management is the core challenge (e.g. cloud) or the online construction and evolution are key enablement for business (e.g. internet of things or cyber-physical-systems).

3.18 EAGLE: Engineering softwAre in the ubiquitous Globe by Leveraging uncErtainty

Paola Inverardi (Universitá di L'Aquila, IT)

Joint work of Inverardi, Paola; Marco Autili; Vittorio Cortellessa; Davide Di Ruscio; Patrizio Pelliccione; Massimo Tivoli
Main reference M. Autili, V. Cortellessa, D. Di Ruscio, P. Inverardi, P. Pelliccione, M. Tivoli, "EAGLE: engineering software in the ubiquitous globe by leveraging uncErtainty," in Proc. of 19th ACM SIGSOFT Symp./13th European Conf. on Foundations of Software Engineering (ESEC/FSE '11), pp. 488–491, 2011.
URL http://dx.doi.org/10.1145/2025113.2025199

In the next future we will be surrounded by a virtually infinite number of software applications that provide computational software resources in the open Globe. Users will be keen on producing their own piece of software, by also reusing existing software, to better satisfy their needs, therefore with a goal oriented, opportunistic use in mind. The produced software will need to be able to evolve, react and adapt to a continuously changing environment, while guaranteeing dependability. The strongest adversary to this view is the lack of knowledge on the software's structure, behaviour, and execution context. Despite the possibility to extract observational models from existing software, a producer will always operate with software artefacts that exhibit a degree of uncertainty in terms of their functional and non functional characteristics. We believe that uncertainty can only be controlled by making it explicit and by using it to drive the production process itself. In this paper, we introduce a novel paradigm of software production process that explores available software and assesses its degree of uncertainty in relation to the opportunistic goal G, assists the producer in creating the appropriate integration means towards G, and validates the quality of the integrated system with respect to G and the current context.

3.19 Models@RT to support Interoperability

Valerie Issarny (INRIA Le Chesnay, FR)

Interoperability is a fundamental challenge for today's extreme distributed systems.Indeed, the high-level of heterogeneity in both the application layer and the underlying middleware and the conflicting assumptions that each system makes about its running environment hinder the successful interoperation of independently-developed systems. Many solutions that aggregate the disparate systems in a non-intrusive way have been proposed.

These solutions use intermediary software entities, called mediators, to interconnect systems despite disparities in their data and/or interaction models by performing the necessary coordination and translations while keeping them loosely-coupled. Creating mediators requires a substantial development effort and a thorough knowledge of the application-domain, which is best understood by domain experts. Moreover, the increasing complexity of today's distributed systems, sometimes referred to as Systems of Systems, makes it almost impossible to develop 'correct' mediators manually. Therefore, formal approaches are used to automatically synthesize mediators. The notion of mediator is further realized using emergent middleware.

In this context, run-time models are used to capture meta-information about the heterogeneous systems including their interfaces and associated behaviour.

This is supplemented by ontological information to enable semantic interoperability in given application domains. We examine the nature of such run-time models coupled with consideration of the supportive synthesis algorithms that use these models to generate the appropriate mediators in order ensure interoperability in highly-heterogeneous environments.

This work takes place in the CONNECT project, which also examines how the models are derived, and how the system can adapt to underlying changes in context or issues related to the performance or behaviour of the system.

3.20 Hyper-Agility: Handling Variability from Design-Time to Runtime

Jean-Marc Jezequel (IRISA – Rennes, FR)

Hyper-agility can be defined as the transposition of the Agile Manifesto at runtime to obtain systems able to adapt automatically to changes in their environment or their user requirements. We present an operational approach based on the use of models to separate concerns by abstracting specific aspects of reality. This approach has become quite popular in recent years for software analysis and design, relying on modelling languages of the UML family. Of course, the separation of concerns is of limited value if we cannot automatically reconstruct these concerns. The automatic composition of models makes it possible to effectively manage changes in the design or maintenance of software, especially in the context of product lines engineering. Beyond the resolution of this issue in the design phase, we show how the composition of models can also be used during the execution of a system to specify and manage dynamically adaptive software systems, here conceptualized as dynamic software product lines.

3.21 Multi-model adaptive control

Marin Litoiu (York University – Toronto, CA)

An adaptive feedback loop in the cloud governs how and when resources (e.g., application server instances) are added to and/or removed from a cloud environment. The adaptive feedback loop can be implemented as a conventional control loop or as a set of heuristic rules.

In the control-theoretic approach, complex constructs such as tracking filters, estimators, regulators, and controllers are utilized. In the heuristic, rule-based approach, various alerts (e.g., events) are defined on instance metrics (e.g., CPU utilization), which are then aggregated at a global scale in order to make provisioning decisions for a given application tier.

This work provides an overview of our experiences designing and working with both approaches to construct an auto-scaler for simple applications. We enumerate different criteria such as design complexity, ease of comprehension, and maintenance upon which we form an informal comparison between the different methods. We conclude with a brief discussion of how these approaches can be used in the governance of resources to better meet a high-level goal over time.

3.22 Including humans in the Models@Run.Time loop

Brice Morin (SINTEF – Oslo, NO)

Software systems are becoming more and more adaptive. While some software systems operate in a fully autonomic mode (typically, embedded systems), we foresee that the end-users have an important role to play in the new emerging adaptive systems such as Cyber-physical systems or Ambient Assisted Living systems. Such systems will still keep an certain part of autonomy, but their behaviour (core logic and adaptation logic) should be open, to some extent, to customization and personalization so that they can more easily be accepted by end-users. At design-time, designers only have a rough idea on how the system should behave and adapt, however the real requirements and needs will only be discovered at runtime, when actual end-users come into play. We believe that models@runtime could offer i) the right abstractions to end users to customize adaptive systems to their actual needs, and ii) assurance that their decisions will only be enacted if they do not jeopardize their (or the system's) safety.

3.23 Requirements for Models at Runtime

Pieter J. Mosterman (The MathWorks Inc. – Natick, US)

Model-Based Design has become critical to success in industry where embedded systems play an important role in the final products. For example, in automotive industry, the Chevy Volt includes over a million lines of code, which to a large extent have been automatically generated from models. While in a tightly controlled process of system design, models have proven themselves of unrivalled value, the requirements for the use of models at runtime are distinctly different. For example, models to be used for runtime adaptation must be defined for a much broader set of assumptions and these assumptions must be captured explicitly and precisely. Another example is the necessity to have models of potential system adaptations after deployment. The study of the idiosyncrasies that runtime use of models impose are the main objective of this work.

3.24 Models@Runtime for Self-Adaptation an Self-Protection

Liliana Pasquale (University of Limerick, IE)

M@RT make possible to engineer systems that may adapt their behaviour in response to changes in the environment or in the requirements they are supposed to meet. As a matter of facts despite adaptation capabilities could be modelled with great detail at design time, anticipating all possible adaptations is not always feasible. To address this problem the requirements model of the system, which also includes the adaptation capabilities, is conceived as a run- time entity. My research applies requirements@runtime in two different application domains: adaptive service compositions and adaptive security.

For the first application domain, we use a live goal model to represent requirements (including adaptation capabilities) and track their changes. Note that changes in the requirements model cannot be learnt automatically, but must be always planned by the designer. These changes are propagated onto the running service instances (e.g., BPEL processes) through a suitable infrastructure, which is composed of a BPEL engine (to execute application instances), a set of data collectors (to monitor the environment and the applications execution state), monitors (to assess requirements) and adaptors (to modify the running application instances, for example, by changing their service components or by deploying new versions of the process).

On the other end, adaptive security is concerned on dynamically enabling a different set of security countermeasures when assets (to be protected) change unexpectedly, new threats arise, or undiscovered vulnerabilities are revealed. In our approach we relate the asset model to the requirements of the system, which are expressed through a goal model, and the objectives of an attacker, which are expressed through an anti-model. The asset, goals and threat models are conceived as a runtime entity and are used as input to build a causal network to analyse system security in different situations, and enable, when necessary, a set of countermeasures to mitigate the security threat. The relevant modifications that may arise at runtime (e.g., new or changing assets/threats/vulnerabilities) are dynamically discovered and tracked in the three models and consequently change the actual set of countermeasures that are enabled at runtime.

3.25 Automatic Synthesis of Behaviour Protocols for Composable Web-Services

Patrizio Pelliccione (Univ. degli Studi di L'Aquila, IT)

Main reference A. Bertolino, P. Inverardi, P. Pelliccione, M. Tivoli, "Automatic synthesis of behavior protocols for composable web-services," in Proc. of 7th Joint Meeting of the Europ. Software Engineering Conference and the ACM SIGSOFT Symp. on The Foundations of Software Engineering (ESEC/FSE '09), pp. 141–150, 2009.
URL http://dx.doi.org/10.1145/1595696.1595719

Web-services are broadly considered as an effective means to achieve interoperability between heterogeneous parties of a business process and offer an open platform for developing new composite web-services out of existing ones. In the literature many approaches have been proposed with the aim to automatically compose web-services. All of them assume that, along

with the web-service signature, some information is provided about how clients interacting with the web-service should behave when invoking it. We call this piece of information the web-service behaviour protocol.

Unfortunately, in the practice this assumption turns out to be unfounded.

To address this need, in this paper we propose a method to automatically derive from the web-service signature an automaton modelling its behaviour protocol. The method, called StrawBerry, combines synthesis and testing techniques. In particular, synthesis is based on data type analysis. The conformance between the synthesized automaton and the implementation of the corresponding web-service is checked by means of testing. The application of StrawBerry to the Amazon E-Commerce Service shows that it is practical and realistic.

3.26 Continuous Requirements Engineering for Self-Adaptive Software Systems

Anna Perini (CIT- FBK – Povo, IT) and Nauman Ahmed Qureshi (Fondazione Bruno Kessler – Trento, IT)

Joint work of Perini, Anna; Qureshi A. Nauman
Main reference N.A. Qureshi, I. Jureta, A. Perini, "Requirements engineering for self-adaptive systems: Core ontology and problem statement," in 23rd Int'l. Conf. on Advanced Information Systems Engineering (CAiSE'11), Vol. 6741 of LNCS, pp. 33–47, Springer, 2011.
URL http://dx.doi.org/10.1007/978-3-642-21640-4_5

We recently proposed a formulation for the Requirements Problem of SAS (revisiting foundational work by Zave et. al., TOSEM'97, and a more recent one by Jureta et al., RE'08), as a dynamic problem (Qureshi at al., CAISE'11).

According to our formulation elements that contribute to the problem definition can change at run-time, for example, domain assumptions, context, resources, user's preferences.

In our approach the requirements problem is represented through a goal- oriented model, which needs to be dynamically updated in order to reflect run- time changes in the requirements problem.

This motivates our interest in M@RT, and especially in continuous re-appraisal of requirements at run-time, which calls for effective and light-way methods for model representation and reasoning at run-time.

3.27 A Model-based Framework for Dynamic Adaptive Systems

Andres J. Ramirez (Michigan State University, US)

Joint work of Ramirez, Andres J.; Cheng, Betty H.C.

A dynamically adaptive system (DAS) observes itself and its execution environment at run time to detect conditions that warrant adaptation. If an adaptation is necessary, then a DAS changes its structure and behaviour in order to continuously satisfy its requirements, even as its environment changes.

However, it is a challenging task to systematically and rigorously develop a DAS due to environmental uncertainty. In particular, it is often infeasible for a human to identify or enumerate all possible combinations of environmental conditions that a DAS might encounter throughout its lifetime. Nevertheless, a DAS must continuously satisfy its requirements despite the threat that this environmental uncertainty poses to its adaptation capabilities. For my dissertation research I have proposed a model-based framework that supports the specification, monitoring, and dynamic reconfiguration of a DAS to explicitly address uncertainty. Specifically, the proposed framework uses a goal-oriented requirements model to derive utility functions for requirements monitoring in a DAS. Using these functions, our framework then harnesses evolutionary computation techniques to identify interesting combinations of environmental conditions that may adversely affect the behaviour of a DAS and generates adaptations on-demand that not only transition the DAS to a target system configuration, but also preserve system consistency. We have demonstrated the capabilities of our model-based framework by applying it to two different case studies, one that involves an intelligent vehicle system that performs adaptive cruise control, lane keeping, and collision avoidance features, and another that involves the reconfiguration of a network of remote data mirrors.

3.28 The need of M@RT in Event-driven Process-centric Decision Support

David Redlich (Lancaster University, GB)

Joint work of Redlich, David; Wasif Gilani
Main reference D. Redlich, W. Gilani, "Event-driven Process-centric Performance Prediction via Simulation," in Business Process Management Workshops (BPM'11) Part I, vol. 99 of LNBIP, pp. 473–478, Springer, 2011.
URL http://dx.doi.org/10.1007/978-3-642-28108-2_46

Today's fast, competitive and extremely volatile markets exert a great deal of pressure on businesses to react quicker against the changes, and sometimes even before the changes actually happen. A late action can potentially result in a legal compliance failure or violation of service level agreements (SLA's). A business analyst needs to be notified before these failures and violations occur. We propose a simple approach that enables real-time and process-centric decision support in the form of performance prediction with the help of performance models at run time. To achieve this the ability of simulations to produce future-events, which are of the same type like the live-events generated by the really executed business process, is utilised. Live-events and simulated future-events can therefore be treated by a Complex-Event Processing (CEP) engine in the same way and parameter models representing the historic, current, and future performance of the business process can be easily computed or adapted.

Furthermore, we discussed further functional enhancements of the proposed architecture to support, for instance, automated business process management through self-adaptation.

3.29 Delta Modelling for Architectural Reconfiguration at Runtime

Bernhard Rumpe (RWTH Aachen, DE)

Joint work of Haber, Arne; Rumpe, Bernhard; Schäfer, Ina

Given that logically or physically distributed systems such as cars, internet, clouds with attached mobiles etc. need to be designed, modelling languages for distributed architectures are inevitable. One such language, called MontiArc, resembles distributed communication of hierarchically decomposed components over explicit synchronous or asynchronous communication lines in the spirit of Automotive's function nets or UML's communication diagrams.

To describe variability of product lines in architectures, we extended this language for explicit modelling of deltas: structural changes of the architecture that extend functionality, signature, etc. but also allows to restructure communication lines.

So far, deltas have mainly been applied at design time to cut a concrete product from the possible configurations.

In this talk, we discuss how the Delta Modelling approach can be adapted to describe runtime re-configuration of software architectures.

While the modelling language basically remains the same, the underlying technical infrastructure, the way how to configure the architecture and in particular the dynamic re-configuration of the system during runtime significantly change.

3.30 Run-time Model Projections for Software Failure Prediction

Giordano Tamburrelli (Politecnico di Milano, IT)

Software is the driving engine of modern society. Most human activities—including critical ones—are either software enabled or entirely managed by software. As software is becoming ubiquitous and society increasingly relies on it, the adverse impact of unreliable or unpredictable software cannot be tolerated. Indeed, software systems must be able to evolve accordingly to their deployment environment to guarantee a seamless fulfilment of desired requirements and obtain a minimum downtime. In response to this challenge, current Software Engineering aims at designing self-adaptive systems able to react and reconfigure themselves minimizing human intervention and ideally guaranteeing a lifelong requirement fulfilment. To date, Software Engineering research in self-adaptive systems has produced promising initial results.

However, even if these findings provide an essential step towards a set of effective and efficient solutions for self-adaptation building these dependable systems is still unclear and requires further investigation. The most promising approaches to software self-adaptation rely on the introduction at run-time of software models (e.g., Markov models, automata, queuing networks, etc.), which represent the behaviour of the system under scrutiny. Such models are used to analyse and control the running implementation of the system with respect to desired requirements.

However, most current research efforts focus on enabling self-adaptation once the root event which triggered the adaptation already occurred. As a consequence, the system may

have been already compromised with a requirement violation. Designers must ensure that any critical requirement of the system continue to be satisfied before, during and after unforeseen changes. In current software engineering paradigms, systems do not anticipate events which may lead to failures, but only react accordingly to them. Let us consider for example the case in which a change itself is directly connected to a system requirement. In such scenarios the system requirement has been already violated because the occurred change or deviation implies a violation of the requirement. In this case self-adaptation boils down to a self-healing countermeasure to the occurred failure (i.e., reactive adaptation). Current approaches focus on this kind of adaptation and thus may occur into failures even if they adopt the proper countermeasures. They do not anticipate potentially dangerous changes. Run-time failure prediction, in our vision, is the enabling factor to tackle this problem.

Our current research challenge relies on the introduction of the novel concept of Run-Time Model Projection for Failure Prediction. By this we mean the ability of a software system at run-time to automatically forecast potentially dangerous events by reasoning on models which represent the expected future behaviour of the system (i.e., models projections) and thus work around predicted failures before their occurrence. This approach empowers self-adaptation capabilities of software systems obtaining an increased degree of dependability and availability.

Run-Time Model Projection for Failure Prediction is the ability of a system to anticipate failures by: (1) collecting at run-time relevant events occurred in the system internals as well as in its deployment environment, (2) performing quantitative and qualitative analyses of such events with their associated trends, and (3) building models representing a projection of the system behaviour in its near future. Indeed, by collecting these pieces of information and by building model projections it is possible to reason about the future compliance among the system and its requirement exploiting model-checking at run-time.

By analysing this compliance it is possible to predict future failures and put in action adaptation strategies that anticipate the incoming failure and work around it. The long-term vision of this approach is systems that are able to: (1) reason about potential changes, (2) anticipate them, and (2) reason on potential adaptations they might make. This process is driven by the events recorded in the internals of the system plus its environment and the account is given in terms of system requirements through model-checking.

We aim at contributing to the research on models at runtime for self-adaptive systems by building software models at run-time which do not represent the current behaviour of the system but its projection in the immediate future. Given these models, it is possible to automatically reason about the future fulfilment of desired requirements and potentially trigger proper adaptation strategies on time to steer the system and avoid a predicted failure. We investigate quantitative and qualitative algorithms which detect, record, correlate and analyse events occurred in the system and its environment. By reasoning on such events we estimate the future value of specific properties of the system (e.g., response time, incoming requests per unit of time, etc.) and synthesizes such estimates as updates for a software model of the system (initially produced at design-time). Indeed, we update the models at run-time coherently with these estimates obtaining model projections. By this we mean software models that aim at representing a prediction of the system in its near future. By reasoning on model projections it is possible to detect incoming failures before their occurrence. The research challenge tackled in this step relies on the accuracy and reliability of the model projections that is obtained with ad hoc techniques mainly based on statistical algorithms.

Furthermore, it is crucial in run-time model based approaches to have an effective and efficient reasoning engine. We plan to adopt as reasoning engines model-checking

tools, such as PRISM, which will bi able to automatically prove the compliance of model projections against requirements. The introduction at run-time of model-checking is a promising approach for self-adaptive systems and offers several advantages. First of all, it allows a direct mapping of desired requirements against the model of the system. By this we mean that model checking allow not only to detect a requirement violation but also produce insights concerning the originating cause and about possible reconfigurations of the system which may solve the violation through what-if analyses. Secondly, model checking techniques rely on well-known and effective algorithms that are available off-shelf as powerful and open-source tools. Moreover, the recent research efforts on optimizing their performance in terms of execution time and space allow an efficient adoption at run-time, even if there is room for improvement.

3.31 Models@run.time in self-* systems

Matthias Tichy (Universität Paderborn, DE)

Using models@run.time is a key element to develop self-* applications. They can be used to assess the state of the system, to forecast system behaviour at runtime and to adapt system accordingly. We present applications of models@run.time in the area of self-healing distributed systems using graphs and graph transformations, self-adaptation of performance critical business information systems using performance annotated component models, self-adaptation in safety-critical embedded systems using failure propagation models and graph transformations and control engineering using models of the physics at run-time. Finally, we discuss the problem of ensuring the consistency of multiple models and dealing with uncertainty as key challenges in developing systems with models@run.time.

3.32 Safety Models@Run-Time in Open Adaptive Systems

Mario Trapp and Daniel Schneider (Fraunhofer IESE – Kaiserslautern, DE)

Joint work of Schneider, Daniel; Trapp, Mario
Main reference D. Schneider, M. Trapp, "A Safety Engineering Framework for Open Adaptive Systems," in Proc. of the 5th IEEE Int'l Conf. on Self-Adaptive and Self-Organizing Systems, pp. 89–98, IEEE CS, 2011.
URL http://dx.doi.org/10.1109/SASO.2011.20

In embedded systems there is a clear trend towards open, context adaptive systems like cyber physical systems, ubiquitous computing, or ambient intelligence. All of these types of systems share the characteristic that neither their structure/behaviour nor the environment they run in are completely predictable at design time. On the one hand, this means that the systems must be able to adapt to these changing contexts at run time. On the other hand, this means that it is hardly possible to perform a complete quality assurance at design time. Particularly in the context of safety-critical systems this leads to new challenges. Models at run time provide a means to overcome this challenge. They enable the system to reason about its quality at run time and to adapt accordingly to assure the safety of the system in any given context situation. Regarding traditional safety engineering, there is a clear trend

to model-driven safety engineering and to modular certification based on models at design time. Combining these approaches with models at runtime to "safety models at runtime (SM@RT)" it is possible to shift certain safety engineering tasks to run time without losing the control over the systems' quality. Particularly, contracts based on safety certification models at run time seem to provide a very efficient means to ensure safety in open adaptive systems. In general, ensuring quality in adaptive systems poses several research challenges, M@RT build a sound basis to overcome these challenges.

3.33 A generalized view on models@run.time

Frank Trollmann (TU Berlin, DE)

The use of models@run.time is a suitable technique for the implementation of complex self-adaptive systems. The causal connection between the model and its system under study plays an important role in this approach. Although models@run.time are defined in to the scope of software engineering there are other systems which also utilize the causal connection. A comparison of models@run.time and such other approaches could lead to interesting new insights. For this purpose we introduce an abstracted understanding of models@run.time that can also be applied outside the scope of software engineering and makes these approaches comparable in the form of a mega model.

3.34 SmarterContext: Realizing Dynamic Context Management by Exploiting Context Models@Runtime

Norha Milena Villegas and Hausi Müller (University of Victoria, CA)

Self-adaptive software systems adapt their behaviour to address changes in functional and non-functional requirements according to environmental conditions. Over the past decade the dynamic capabilities of these systems have proliferated and improved significantly. However, their real application is still limited due to a lack of mechanisms to address the uncertainty of the execution environment. The execution environment of an adaptive system is composed of external and internal context entities that can affect the desired system's behaviour. Thus, these entities and the interactions among them must be monitored to support decision making in the adaptation process. Nevertheless, as context information evolves over time, relevant entities and the corresponding monitoring requirements cannot be fully specified at design-time. Moreover, as the system is also evolving, monitoring requirements continuously change accordingly. Therefore, the effectiveness of self-adaptation is completely dependent on the adaptive capability of monitoring mechanisms to preserve context-awareness throughout the adaptation process. SmarterContext—our innovative approach to dynamic context management—exploits adaptive software techniques and RDF-based context models at runtime to optimize context- awareness and self-adaptation in service-oriented software systems.

SmarterContext is equipped with a self-adaptive context management infrastructure and an extensible context taxonomy based on the resource description framework (RDF) in order to support dynamic changes in context management strategies, through the deployment of new context management components to keep track of changes in relevant context at run-time. Our SmarterContext taxonomy includes a set of inference rules for supporting dynamic context representation and reasoning.

3.35 Models at Runtime for Adaptive and Self-managing Software

Thomas Vogel (Hasso-Plattner-Institut – Potsdam, DE)

Software systems have to be continuously adapted to their changing requirements or environments. This is typically done by maintenance, while self-aware, context-aware, ultra-large-scale, or mission-critical software systems often have to be adapted during operation (self-adaptive software). Model-driven engineering and models at runtime play crucial roles in supporting maintenance, enabling self-adaptation, and integrating both, maintenance and self-adaptation.

In our talk, we outline our perspective on "models at runtime" as any models used on-line, i.e., internal to the running system, to represent running software, to represent the software's environment, or to manipulate or analyse any of the former two. Moreover, we sketch the benefits of models at runtime for maintenance, self-adaptation, and for an integrated setting of maintenance and self-adaptation. Considering different kinds of runtime models required for self-adaptation and self-adaptation activities as model operations, we discuss the role of a megamodel for specifying adaptation loops for self-adaptive software and operationalising the treatment of runtime models as means for monitoring, analysis, planning, and executing planned changes.

4 Working Groups

This section lists the abstracts of the break-out sessions. Each break-out group is in charge of a main section of the roadmap that is under preparation.

4.1 Group 1: Assurance Using Models@Run.Time for Self-Adaptive Systems

Betty H. C. Cheng, Kerstin I. Eder, Martin Gogolla, Lars Grunske, Marin Litoiu, Hausi A. Müller, Patrizio Pelliccione, Anna Perini, Nauman A. Qureshi, Bernhard Rumpe, Daniel Schneider, Frank Trollmann, Norha M. Villegas

Traditionally, software assurance is part of the software development process; i.e. assurance is a development-time concern. For software systems that change or evolve at runtime, as

in self-adaptive and self-managing systems, however, software assurance becomes a critical runtime concern. Continuous assurance throughout the entire software life-cycle provides unprecedented opportunities for monitoring, analyzing, guaranteeing, and predicting system properties and qualities throughout the operation of a software system. The fact that many variables that are free at design-time are bound at runtime provides new opportunities for verification and validation at runtime leading to assurance of critical system properties.

The goal of the assurance breakout group of Dagstuhl Seminar 11481 on Models@runtime was to identify research questions for software assurance using models at runtime to advance the state of the art in runtime assurance. Initially a set of valuable background starting points was established. The following is a summary of the most important research questions from the assurance breakout group:

- What development-time assurance methods and techniques (i.e., the entire spectrum from light-weight to heavy-weight approaches) and models (i.e., descriptive, prescriptive, constructive and predictive) readily extend to runtime?
- How do traditional assurance models and methods from domains such as performance, safety, and reliability extend to runtime?
- How can we extend reference models for self-adaptive and self-managing systems to include models and assurance at runtime (e.g., MAPE-K loop)?
- Can we leverage runtime assurance techniques from other disciplines (e.g., control theory)?
- Assuming models at runtime what reference architectures are appropriate for assurance at runtime?
- What are appropriate assurance reasoning techniques for different phases of the software life cycle (i.e., development, installation, load, and runtime) and how can assurance results from different life-cycle phases be combined to assure systems at runtime (e.g., incremental and compositional assurance)?
- How can we partition runtime assurance according the different types of runtime changes (e.g., dynamic context, changing requirements, or evolving models)?
- What are ideal applications to demonstrate the challenges and opportunities for assurance using models at runtime?

The research questions on runtime assurance will become part of a wider research roadmap dedicated to the development of models at runtime. One key challenge for this entire field is to develop effective training methods that equip traditional software engineers with the knowledge and skills to design and build software systems that change or evolve at runtime.

4.2 Group 2: Runtime Updating/Adaptation Mechanisms

Amel Bennaceur, Mehmet Aksit, Robert France, Walter Cazzola, Fabio M. Costa, Pär Emmanuelson, Nikolaos Georgantas, Huang Gang, Pieter J. Mosterman, David Redlich, Alfonso Pierantonio, Giordano Tamburrelli, Thomas Vogel and Matthias Tichy

Runtime adaptation is becoming a fundamental property of today's complex open systems. Models at runtime (M@RT) systems offer promising abstractions, techniques and tools to manage the increasing complexity of adaptation and to handle the high dynamism of these systems. In this group, we discussed the issues and the solutions for runtime adaptation across different domains (e.g., fault-tolerance systems, dynamic tuning, energy optimization).

Despite the recent efforts from the software engineering community on M@RT, there is a lack of techniques and methods for the proper integration across domains, across layers (e.g., architectural, middleware, code), and across concerns (e.g., performance, security, reliability). We agreed on a conceptual architecture for adapting systems using M@RT and identified the different mechanisms used to: (1) create or update the system model according to the evolution of the system or to changes in its environment, (2) automatically reason on M@RT to produce appropriate adaptation strategies, (3) analyse and maintain different M@RT, each representing a specific view and/or abstraction, and (4) propagate changes from the model back to the system. Finally, we considered each mechanism, reviewed related work, and identified the open challenges in the form of short-term and long-term research goals. The output of this work leads to a fascinating set of research challenges both in terms of understanding the characteristics of M@RT and using them to create appropriate adaptation mechanisms in open settings, especially for the Internet of Things and Cyber-Physical systems. This is a vast and largely unchartered territory and we invite other researchers to join in the quest for suitable solutions for models at runtime.

4.3 Group 3: Uses and Purposes of M@RT Systems

Uwe Aßmann, Sebastian Götz, Jean-Marc Jezequel, Brice Morin and Mario Trapp

The goal of this section is to understand how models@runtime are key enablers for modern software systems, to clarify their typical use cases and fundamental interests. Traditionally the development of software systems used to be split in distinct steps with a clear distinction between design activities and runtime execution. Safety critical embedded systems are for example designed and intensively validated at design time (e.g., using model checking) before they are actually deployed. At runtime, they have a predictable behaviour, time and resource consumption, which make it possible for certification bodies to approve these systems. The development of adaptive systems requires negotiating a trade off between adaptiveness and predictability. In many cases, system adaptation is restricted to a set of well-defined modes and reconfigurations defined at design-time. However, emerging types of systems—for example, cyber-physical systems—may need to adapt in ways that cannot be foreseen at design time and therefore require a higher degree of runtime adaptability, which hinders predictability. Hence, new approaches are needed to enable unanticipated adaptations while ensuring guarantees: This is, in our opinion, the ultimate purpose of models@runtime. The key characteristic of a M@RT system is its ability to project some aspects of the reality (e.g., its context, its behaviour, its goals) to the modelling space in order to enable tractable decision-making that produces decidable plans. This is basically separation of concerns applied in a disciplined way at runtime, and to some extent, is analogous to how human thinking works. This enables systems to reason about alternatives to reach their goals and to determine consequences of particular actions, rather than just learn and react (i.e., display animal-like behaviour) as is done in classical systems. This requires that the system be able to justify why a certain decision was made or not. The key advantage of M@RT systems over reflective software systems is decidability and tractability, which can reconcile users, domain experts, engineers (aware of the obvious need for runtime adaptivity), with certification bodies (which need stringent guarantees). Based on our recent experiences, we propose a

generic Reference Architecture (RA), which can be instantiated to cope with the specificities of different domains.

4.4 Group 4: Living With Uncertainty In the Age of Runtime Models

Nelly Bencomo, Amit K. Chopra, Siobhan Clarke, Holger Giese, Paolo Inverardi, Valerie Issarny, Liliana Pasquale, and Andres J. Ramirez

Systems that operate in dynamic environments are inherently executing in situations of uncertainty. Uncertainty arises from multiple sources, such as the environment, the system itself, and different stakeholders, for example, end users, society including regulatory bodies, and others. Uncertainty emerges when there is a difference between useful information that exists in, for example, the original system and information available in the runtime model at a certain point in time. A runtime model provides an abstraction of information of interest to and in the executing system. It provides a knowledge base for decision-making related to behaviour optimized as far as possible within the bounds of some uncertainty. A runtime model to represent the uncertainty of interest can be specified at design time according to the possible goals of the system. At runtime, a monitor and analyser will instrument and refine information in the runtime model, which may reduce the level of uncertainty. A planner and executor must be able to deal with remaining uncertainty in a robust manner. In this group, we discussed research questions relating to the treatment of uncertainty with models at runtime.

Participants

- Mehmet Aksit
University of Twente, NL
- Joe Armstrong
ERICSSON – Stockholm, SE
- Uwe Aßmann
TU Dresden, DE
- Nelly Bencomo
INRIA, FR
- Amel Bennaceur
INRIA, FR
- Walter Cazzola
Università di Milano, IT
- Betty H. C. Cheng
Michigan State University, US
- Amit K. Chopra
University of Trento – Povo, IT
- Siobhán Clarke
Trinity College – Dublin, IE
- Fabio M. Costa
Univ. Federal de Goiás, BR
- Kerstin I. Eder
University of Bristol, GB
- Pär Emanuelsson
Ericsson AB – Linköping, SE
- Robert B. France
Colorado State University, US
- Nikolaos Georgantas
INRIA Le Chesnay, FR
- Holger Giese
Hasso-Plattner-Institut – Potsdam, DE
- Sebastian Götz
TU Dresden, DE
- Martin Gogolla
Universität Bremen, DE
- Lars Grunske
TU Kaiserslautern, DE
- Gang Huang
Peking University, CN
- Paola Inverardi
Universitá di L'Aquila, IT
- Valerie Issarny
INRIA Le Chesnay, FR
- Jean-Marc Jezequel
IRISA – Rennes, FR
- Marin Litoiu
York University – Toronto, CA
- Brice Morin
SINTEF – Oslo, NO
- Pieter J. Mosterman
The MathWorks Inc. – Natick. US
- Hausi A. Müller
University of Victoria, CA
- Liliana Pasquale
University of Limerick, IE
- Pankesh Patel
INRIA, FR
- Patrizio Pelliccione
Univ. degli Studi di L'Aquila, IT
- Anna Perini
CIT- FBK – Povo, IT
- Alfonso Pierantonio
Univ. degli Studi di L'Aquila, IT
- Nauman Ahmed Qureshi
Fondazione Bruno Kessler – Trento, IT
- Andres J. Ramirez
Michigan State University, US
- David Redlich
Lancaster University, GB
- Bernhard Rumpe
RWTH Aachen, DE
- Daniel Schneider
Fraunhofer IESE – Kaiserslautern, DE
- Giordano Tamburrelli
Politecnico di Milano, IT
- Matthias Tichy
Universität Paderborn, DE
- Mario Trapp
Fraunhofer IESE – Kaiserslautern, DE
- Frank Trollmann
TU Berlin, DE
- Norha Milena Villegas
University of Victoria, CA
- Thomas Vogel
Hasso-Plattner-Institut – Potsdam, DE